Introduction to Probability

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 308

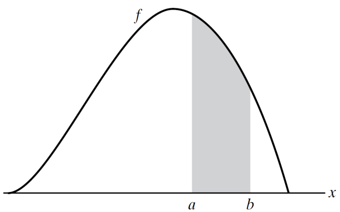
Subsection 1

Introduction

Probability Density Function

- Some random variables have an uncountable set of possible values.
- Two examples are:
  - The time that a train arrives at a specified stop;
  - The lifetime of a transistor.
- Let X be such a random variable.
- We say that X is a **continuous random variable** if there exists a nonnegative function f, defined for all real $x \in (-\infty, \infty)$, having the property that, for any set B of real numbers,

$$P\{X \in B\} = \int_B f(x)dx.$$



$P(a \leq X \leq b)$ = area of shaded region

- The function f is called the **probability density function** of the random variable X.

Properties

- Since X must assume some value, f must satisfy

$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)dx.$$

- All probability statements about X can be answered in terms of f.
- For instance, from, letting $B = [a, b]$, we obtain

$$P\{a \leq X \leq b\} = \int_a^b f(x)dx.$$

- If we let $a = b$, we get $P\{X = a\} = \int_a^a f(x)dx = 0$.
- Thus, the probability that a continuous random variable will assume any fixed value is zero.
- Hence, for a continuous random variable,

$$P\{X < a\} = P\{X \leq a\} = F(a) = \int_{-\infty}^{a} f(x)dx.$$

Example

- Suppose that X is a continuous random variable whose probability density function is given by

$$f(x) = \begin{cases} C(4x - 2x^2), & 0 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

(a) What is the value of C?
(b) Find $P\{X > 1\}$.

(a) Since f is a probability density, we must have $\int_{-\infty}^{\infty} f(x)dx = 1$. This gives

$$C \int_0^2 (4x - 2x^2)dx = 1 \Rightarrow C\left[2x^2 - \frac{2x^3}{3}\right]_{x=0}^{x=2} = 1 \Rightarrow C = \frac{3}{8}.$$

(b) $P\{X > 1\} = \int_1^{\infty} f(x)dx = \frac{3}{8}\int_1^2 (4x - 2x^2)dx = \frac{3}{8}(\frac{8}{3} - \frac{4}{3}) = \frac{1}{2}$.

Example

- The amount of time in hours that a computer functions before breaking down is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} \lambda e^{-x/100}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Find the probability that:

(a) A computer will function between 50 and 150 hours before breaking down;

(b) It will function for fewer than 100 hours.

(a) Note that $1 = \int_{-\infty}^{\infty} f(x)dx = \lambda \int_0^{\infty} e^{-x/100}dx$.

Hence, we get $1 = -\lambda(100)e^{-x/100}|_0^{\infty} = 100\lambda$ or $\lambda = \frac{1}{100}$.

Example (Cont'd)

- Thus, the probability that a computer will function between 50 and 150 hours before breaking down is given by

$$\begin{aligned} P\{50 < X < 150\} &= \int_{50}^{150} \frac{1}{100} e^{-x/100} dx = -e^{-x/100}\Big|_{50}^{150} \\ &= e^{-1/2} - e^{-3/2} \approx 0.384. \end{aligned}$$

(b) Similarly,

$$\begin{aligned} P\{X < 100\} &= \int_{0}^{100} \frac{1}{100} e^{-x/100} dx \\ &= -e^{-x/100}\Big|_{0}^{100} \\ &= 1 - e^{-1} \approx 0.633. \end{aligned}$$

In other words, approximately 63.3 percent of the time, a computer will fail before registering 100 hours of use.

Example

- The lifetime in hours of a certain kind of radio tube is a random variable having a probability density function given by

$$f(x)=\begin{cases} 0, & x \leq 100 \\ \frac{100}{x^2}, & x > 100 \end{cases}$$

What is the probability that exactly 2 of 5 such tubes in a radio set will have to be replaced within the first 150 hours of operation? Assume that the events E_i, $i=1,2,3,4,5$, that the ith such tube will have to be replaced within this time are independent.

From the statement of the problem, we have

$$P(E_i)=\int_0^{150} f(x)dx = 100\int_{100}^{150} x^{-2}dx = 100\left[-\frac{1}{x}\right]_{x=100}^{x=150} = \frac{1}{3}.$$

Hence, from the independence of the events E_i, it follows that the desired probability is $\binom{5}{2}(\frac{1}{3})^2(\frac{2}{3})^3 = \frac{80}{243}$.

Probability Density and Cumulative Distribution

- The relationship between the cumulative distribution F and the probability density f is expressed by

$$F(a) = P\{X \in (-\infty, a]\} = \int_{-\infty}^{a} f(x)dx.$$

  Differentiating both sides yields $\frac{d}{da}F(a) = f(a)$.
  The density is the derivative of the cumulative distribution function.
- A more intuitive interpretation of the density function may be obtained as follows:

$$P\{a - \frac{\varepsilon}{2} \leq X \leq a + \frac{\varepsilon}{2}\} = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f(x)dx \approx \varepsilon f(a),$$

  when ε is small and when $f(\cdot)$ is continuous at $x = a$.
  - The probability that X will be contained in an interval of length ε around the point a is *approximately* $\varepsilon f(a)$.
  - So $f(a)$ is a measure of how likely it is that X will be near a.

Example

- If X is continuous with distribution function F_X and density function f_X, find the density function of $Y = 2X$.

  We will determine f_Y in two ways.

  The first way is to derive, and then differentiate, the distribution function of Y:

$$\begin{aligned} F_Y(a) &= P\{Y \le a\} = P\{2X \le a\} \\ &= P\{X \le \tfrac{a}{2}\} = F_X(\tfrac{a}{2}). \end{aligned}$$

  Differentiation gives

$$f_Y(a) = \frac{1}{2} f_X\left(\frac{a}{2}\right).$$

Example (Cont'd)

- Another way to determine f_Y is to note that

$$\begin{array}{rcl} \epsilon f_Y(a) & \approx & P\{a-\frac{\epsilon}{2} \le Y \le a+\frac{\epsilon}{2}\} \\ & = & P\{a-\frac{\epsilon}{2} \le 2X \le a+\frac{\epsilon}{2}\} \\ & = & P\{\frac{a}{2}-\frac{\epsilon}{4} \le X \le \frac{a}{2}+\frac{\epsilon}{4}\} \\ & \approx & \frac{\epsilon}{2} f_X(\frac{a}{2}). \end{array}$$

Dividing through by ϵ gives the same result as before.

Subsection 2

Expectation and Variance of Continuous Random Variables

Expectation of Continuous Random Variables

- We defined the expected value of a discrete random variable X by

$$E[X] = \sum_x xP\{X = x\}.$$

- If X is a continuous random variable having probability density function $f(x)$, then

$$f(x)dx \approx P\{x \leq X \leq x + dx\} \text{ for } dx \text{ small.}$$

- Hence, the analogous definition is to define the **expected value** of X by

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx.$$

Example

- Find $E[X]$ when the density function of X is

$$f(x) = \begin{cases} 2x, & \text{if } 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

$$E[X] = \int xf(x)dx = \int_0^1 2x^2 dx = \left[\frac{2}{3}x^3\right]_{x=0}^{x=1} = \frac{2}{3}.$$

Example

- The density function of X is given by

$$f(x) = \begin{cases} 1, & \text{if } 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Find $E[e^X]$.
Let $Y = e^X$. We first determine F_Y, the probability distribution function of Y. For $1 \leq x \leq e$,

$$\begin{aligned} F_Y(x) &= P\{Y \leq x\} = P\{e^X \leq x\} = P\{X \leq \log(x)\} \\ &= \int_0^{\log(x)} f(y)dy = [y]_{y=0}^{y=\log x} = \log(x). \end{aligned}$$

By differentiating $F_Y(x)$, we can conclude that the probability density function of Y is given by $f_Y(x) = \frac{1}{x}$, $1 \leq x \leq e$.
Hence,

$$E[e^X] = E[Y] = \int_{-\infty}^{\infty} x f_Y(x)dx = \int_1^e dx = e - 1.$$

Computing the Expectation of a Function of X

Proposition

If X is a continuous random variable with probability density function $f(x)$, then, for any real-valued function g,

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

- Revisiting the preceding example, we obtain

$$E[e^X] = \int_0^1 e^x dx = e - 1.$$

Proof for the Nonnegative Case

Lemma

For a nonnegative random variable Y,

$$E[Y] = \int_0^\infty P\{Y > y\}dy.$$

- We present a proof when Y is a continuous random variable with probability density function f_Y.

  We have

  $$\int_0^\infty P\{Y > y\}dy = \int_0^\infty \int_y^\infty f_Y(x)dxdy,$$

  where we have used the fact that $P\{Y > y\} = \int_y^\infty f_Y(x)dx$.

  Interchanging the order of integration in the preceding equation yields

  $$\begin{aligned} \int_0^\infty P\{Y > y\}dy &= \int_0^\infty (\int_0^x dy) f_Y(x)dx \\ &= \int_0^\infty x f_Y(x)dx = E[Y]. \end{aligned}$$

Proof for the Nonnegative Case (Cont'd)

- From the lemma, for any function g for which $g(x) \geq 0$,

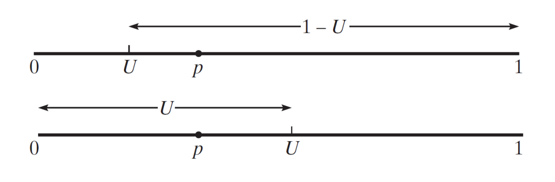
$$\begin{aligned} E[g(X)] &= \int_0^\infty P\{g(X) > y\} dy \\ &= \int_0^\infty \int_{x:g(x)>y} f(x) dx dy \\ &= \int_{x:g(x)>0} \int_0^{g(x)} dy f(x) dx \\ &= \int_{x:g(x)>0} g(x) f(x) dx. \end{aligned}$$

Example

- A stick of length 1 is split at a point U that is uniformly distributed over $(0,1)$.

  Determine the expected length of the piece that contains the point p, $0 \leq p \leq 1$.

  Let $L_p(U)$ denote the length of the substick that contains the point p, and note that



$$L_p(U) = \begin{cases} 1-U, & \text{if } U < p \\ U, & \text{if } U > p \end{cases}$$

Example (Cont'd)

- Hence, from the proposition,

$$\begin{aligned} E[L_p(U)] &= \int_0^1 L_p(u)du \\ &= \int_0^p (1-u)du + \int_p^1 udu \\ &= \left[-\frac{(1-u)^2}{2}\right]_{u=0}^{u=p} + \left[\frac{u^2}{2}\right]_{u=p}^{u=1} \\ &= \frac{1}{2} - \frac{(1-p)^2}{2} + \frac{1}{2} - \frac{p^2}{2} \\ &= \frac{1}{2} + p(1-p). \end{aligned}$$

- Note that $p(1-p) = -p^2 + p$ is maximized when $p = \frac{1}{2}$.

  Thus, the expected length of the substick containing the point p is maximized when p is the midpoint of the original stick.

Example

- Suppose that:
  - being s minutes early for an appointment, incurs the cost cs;
  - being s minutes late, incurs the cost ks.

  Suppose also that the travel time from origin to destination is a continuous random variable having probability density function f.

  Determine the time at which one should depart to minimize the expected cost.

  Let X denote the travel time.

  If departure occurs t minutes before the appointment, then the cost $C_t(X)$ is given by

$$C_t(X) = \begin{cases} c(t - X), & \text{if } X \leq t \\ k(X - t), & \text{if } X \geq t \end{cases}$$

Example (Cont'd)

- Therefore,

$$\begin{array}{rcl} E[C_t(X)] & = & \int_0^\infty C_t(x)f(x)dx \\ & = & \int_0^t c(t-x)f(x)dx + \int_t^\infty k(x-t)f(x)dx \\ & = & ct\int_0^t f(x)dx - c\int_0^t xf(x)dx \\ & & + k\int_t^\infty xf(x)dx - kt\int_t^\infty f(x)dx. \end{array}$$

The value of t that minimizes $E[C_t(X)]$ can now be obtained by calculus. Differentiate and set the derivative equal to zero.

$$\begin{array}{rcl} \frac{d}{dt}E[C_t(X)] & = & ctf(t) + cF(t) - ctf(t) - ktf(t) \\ & & + ktf(t) - k[1-F(t)] \\ & = & (k+c)F(t) - k. \end{array}$$

Thus, the minimal expected cost is obtained when one leaves t^* minutes before the appointment, where t^* satisfies $F(t^*) = \frac{k}{k+c}$.

Linearity of Expectation

Corollary

If a and b are constants, then

$$E[aX+b]=aE[X]+b.$$

- We obtain, using the formula for $E[aX+b]$,

$$\begin{aligned} E[aX+b] &= \int_{-\infty}^{+\infty}(ax+b)f(x)dx \\ &= a\int_{-\infty}^{+\infty}xf(x)dx+b\int_{-\infty}^{+\infty}f(x)dx \\ &= a\cdot E[X]+b\cdot 1 \\ &= aE[X]+b. \end{aligned}$$

Variance of a Continuous Random Variable

- If X is a random variable with expected value μ, then the **variance** of X is defined (for any type of random variable) by

$$\mathrm{Var}(X) = E[(X-\mu)^2].$$

- The alternative formula,

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

is established in a manner similar to its counterpart in the discrete case.

Example

- Find Var(X) for X with density

$$f(x) = \begin{cases} 2x, & \text{if } 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

We first compute $E[X^2]$.

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 2x^3 dx = \left[\frac{x^4}{2}\right]_{x=0}^{x=1} = \frac{1}{2}.$$

We have seen previously that $E[X] = \frac{2}{3}$.

Thus, we obtain

$$\text{Var}(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{18}.$$

Property of Variance

- For constants a and b,

$$\text{Var}(aX+b)=a^2\text{Var}(X).$$

- In fact, using the linearity of the expected value, we get

$$\begin{aligned}
\text{Var}(aX+b) &= E[(aX+b)^2]-E[aX+b]^2 \\
&= E[a^2X^2+2abX+b^2]-(aE[X]+b)^2 \\
&= a^2E[X^2]+2abE[X]+b^2 \\
&\qquad -a^2E[X]^2-2abE[X]-b^2 \\
&= a^2(E[X^2]-E[X]^2) \\
&= a^2\text{Var}(X).
\end{aligned}$$

Subsection 3

The Uniform Random Variable

Uniformly Distributed Random Variables

- A random variable is said to be **uniformly distributed** over the interval $(0,1)$ if its probability density function is given by

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

- This is a density function:
  - For all x, $f(x) \geq 0$;
  - We have

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^1 dx = [x]_{x=0}^{x=1} = 1.$$

Definability and Uniformity

- Note that $f(x) > 0$ only when $x \in (0,1)$.

  This implies that X must assume a value in the interval $(0,1)$.
- Also, $f(x)$ is constant for $x \in (0,1)$.

  Hence, X is just as likely to be near any value in $(0,1)$ as it is to be near any other value.

  To verify this, note that, for any $0 < a < b < 1$,

  $$P\{a \leq X \leq b\} = \int_a^b f(x)dx = b - a.$$

  In other words, the probability that X is in any particular subinterval of $(0,1)$ equals the length of that subinterval.

Uniform Random Variable on (α, β)

- We say that X is a **uniform random variable** on the interval (α, β) if the probability density function of X is given by

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha}, & \text{if } \alpha < x < \beta \\ 0, & \text{otherwise} \end{cases}$$

- Recall that $F(a) = \int_{-\infty}^{a} f(x)dx$.

  Thus, the distribution function of a uniform random variable on the interval (α, β) is given by

$$F(a) = \begin{cases} 0, & \text{if } a \leq \alpha \\ \frac{a-\alpha}{\beta-\alpha}, & \text{if } \alpha < a < \beta \\ 1, & \text{if } a \geq \beta \end{cases}$$

Expectation of Uniform Random Variables

- Let X be uniformly distributed over (α, β).
- Then

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_{\alpha}^{\beta} \frac{x}{\beta-\alpha} dx \\ &= \frac{1}{\beta-\alpha} \left[\frac{x^2}{2} \right]_{x=\alpha}^{x=\beta} \\ &= \frac{\beta^2-\alpha^2}{2(\beta-\alpha)} \\ &= \frac{\beta+\alpha}{2}. \end{aligned}$$

- In words, the expected value of a random variable that is uniformly distributed over some interval is equal to the midpoint of that interval.

Variance of Uniform Random Variables

- Let X be uniformly distributed over (α, β).
- To find $\mathrm{Var}(X)$, we first calculate $E[X^2]$.

$$\begin{aligned} E[X^2] &= \int_\alpha^\beta \frac{1}{\beta-\alpha} x^2 dx \\ &= \frac{\alpha^3-\beta^3}{3(\beta-\alpha)} \\ &= \frac{\beta^2+\alpha\beta+\alpha^2}{3}. \end{aligned}$$

- Hence,

$$\begin{aligned} \mathrm{Var}(X) &= \frac{\beta^2+\alpha\beta+\alpha^2}{3} - \frac{(\alpha+\beta)^2}{4} \\ &= \frac{4\beta^2+4\alpha\beta+4\alpha^2-3\alpha^2-6\alpha\beta-3\beta^2}{12} = \frac{(\beta-\alpha)^2}{12}. \end{aligned}$$

- So the variance of a random variable that is uniformly distributed over (α, β) is the square of the length $\beta - \alpha$ divided by 12.

Example

- If X is uniformly distributed over $(0, 10)$, calculate the probability that:
  (a) $X < 3$;
  (b) $X > 6$;
  (c) $3 < X < 8$.

(a) $P\{X < 3\} = \int_0^3 \frac{1}{10} dx = \frac{3}{10}$.

(b) $P\{X > 6\} = \int_6^{10} \frac{1}{10} dx = \frac{4}{10}$.

(c) $P\{3 < X < 8\} = \int_3^8 \frac{1}{10} dx = \frac{1}{2}$.

Example

- Buses arrive at a specified stop at 15 minute intervals starting at 7 A.M. That is, they arrive at 7, 7:15, 7:30, 7:45, and so on.
  If a passenger arrives at the stop at a time that is uniformly distributed between 7 and 7:30, find the probability that he waits:
  (a) less than 5 minutes for a bus;
  (b) more than 10 minutes for a bus.

  Let X denote the number of minutes past 7 A.M. that the passenger arrives at the stop.

  X is a uniform random variable over the interval $(0, 30)$.

  Thus, the passenger will have to wait less than 5 minutes if (and only if) he arrives between 7:10 and 7:15 or between 7:25 and 7:30.

  Hence, the desired probability for part (a) is

$$P\{10 < X < 15\} + P\{25 < X < 30\} = \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = \frac{1}{3}.$$

Example (Cont'd)

- Similarly, he would have to wait more than 10 minutes if he arrives between 7 and 7:05 or between 7:15 and 7:20.

  So the probability for part (b) is

$$P\{0 < X < 5\} + P\{15 < X < 20\} = \int_0^5 \frac{1}{30} dx + \int_{15}^{20} \frac{1}{30} dx = \frac{1}{3}.$$

Bertrand's Paradox (Geometrical Probability)

- Consider a random chord of a circle.

  What is the probability that the length of the chord will be greater than the side of the equilateral triangle inscribed in that circle?
- As stated, the problem is incapable of solution because it is not clear what is meant by a random chord.
- To give meaning to this phrase, we reformulate the problem in two distinct ways.

Bertrand's Paradox: First Formulation

- The position of the chord can be determined by its distance from the center of the circle.

  This distance can vary between 0 and r, the radius of the circle.

  The length of the chord will be greater than the side of the equilateral triangle inscribed in the circle if the distance from the chord to the center of the circle is less than $\frac{r}{2}$.

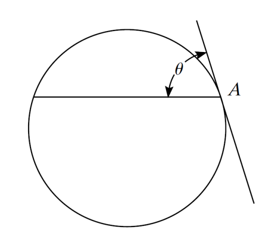
  Assume that a random chord is a chord whose distance D from the center of the circle is uniformly distributed between 0 and r.

  Then the probability that the length of the chord is greater than the side of an inscribed equilateral triangle is

$$P\left\{D < \frac{r}{2}\right\} = \frac{r/2}{r} = \frac{1}{2}.$$

Bertrand's Paradox: Second Formulation

- Consider an arbitrary chord of the circle. Through one end of the chord, draw a tangent. The angle θ between the chord and the tangent can vary from 0° to 180° and determines the position of the chord.



The length of the chord will be greater than the side of the inscribed equilateral triangle if the angle θ is between 60° and 120°.

Assume that a random chord is a chord whose angle θ is uniformly distributed between 0° and 180°.

Then the desired answer in this formulation is

$$P\{60 < \theta < 120\} = \frac{120 - 60}{180} = \frac{1}{3}.$$

Bertrand's Paradox: Comments

- Note that random experiments could be performed in such a way that $\frac{1}{2}$ or $\frac{1}{3}$ would be the correct probability.
  - Assume a circular disk of radius r is thrown on a table ruled with parallel lines a distance $2r$ apart.
    Then one and only one of these lines would form a chord.
    All distances from this chord to the center would be equally likely.
    So the desired probability that the chord's length will be greater than the side of an inscribed equilateral triangle is $\frac{1}{2}$.
  - Assume the experiment consisted of rotating a needle freely about a point A on the edge of the circle.
    In this case, the desired answer would be $\frac{1}{3}$.

Subsection 4

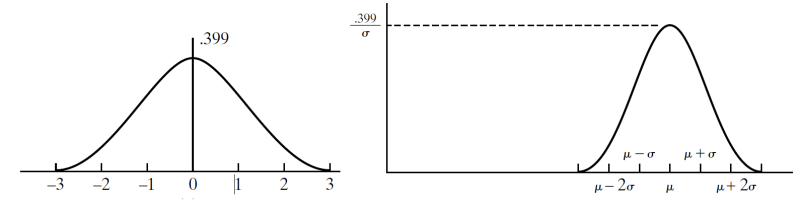
Normal Random Variables

Normal Random Variables

- We say that X is a **normal random variable**, or simply that X is **normally distributed**, with parameters μ and σ^2 if the density of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty.$$

- This density function is a bell-shaped curve that is symmetric about μ.



The Normal is a Probability Density Function

- We show that $f(x)$ is indeed a probability density function, i.e., that

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx = 1.$$

- Making the substitution $y = \frac{x-\mu}{\sigma}$, we see that

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2} dy.$$

- Hence, we must show that $\int_{-\infty}^{\infty} e^{-y^2/2} dy = \sqrt{2\pi}$.
- Toward this end, let $I = \int_{-\infty}^{\infty} e^{-y^2/2} dy$.
- Then

$$I^2 = \int_{-\infty}^{\infty} e^{-y^2/2} dy \int_{-\infty}^{\infty} e^{-x^2/2} dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(y^2+x^2)/2} dy dx.$$

The Normal is a Probability Density Function (Cont'd)

- We now evaluate the double integral by means of a change of variables to polar coordinates ($x = r\cos\theta$, $y = r\sin\theta$ and $dydx = rd\theta dr$).
- Thus,

$$\begin{aligned} I^2 &= \int_0^\infty \int_0^{2\pi} e^{-r^2/2} r d\theta dr \\ &= 2\pi \int_0^\infty r e^{-r^2/2} dr \\ &= -2\pi e^{-r^2/2}\big|_0^\infty = 2\pi. \end{aligned}$$

- Hence, $I = \sqrt{2\pi}$, and the result is proved.

Mean and Variance Transformations

- If X is normally distributed with parameters μ and σ^2, then $Y = aX + b$ is normally distributed with parameters $a\mu + b$ and $a^2\sigma^2$.
- To prove this statement, suppose that $a > 0$.
- Let F_Y denote the cumulative distribution function of Y.
- Then

$$\begin{aligned} F_Y(x) &= P\{Y \leq x\} = P\{aX + b \leq x\} \\ &= P\{X \leq \tfrac{x-b}{a}\} = F_X(\tfrac{x-b}{a}), \end{aligned}$$

  where F_X is the cumulative distribution function of X.
- By differentiation, the density function of Y is then

$$\begin{aligned} f_Y(x) &= \tfrac{1}{a} f_X(\tfrac{x-b}{a}) \\ &= \frac{1}{\sqrt{2\pi} a\sigma} \exp\{-(\tfrac{x-b}{a} - \mu)^2 / 2\sigma^2\} \\ &= \frac{1}{\sqrt{2\pi} a\sigma} \exp\{-(x - b - a\mu)^2 / 2(a\sigma)^2\}. \end{aligned}$$

- This shows that Y is normal with parameters $a\mu + b$ and $a^2\sigma^2$.

Standard Normal Random Variables

- An important implication of the preceding result is that if X is normally distributed with parameters μ and σ^2, then $Z = \frac{X-\mu}{\sigma}$ is normally distributed with parameters 0 and 1.
- Such a random variable is said to be a **standard**, or a **unit**, **normal random variable**.

Expectation and Variance of a Normal Random Variable

- Find $E[X]$ and $\text{Var}(X)$ when X is a normal random variable with parameters μ and σ^2.
- We start by finding the mean and variance of the standard normal random variable $Z = \frac{X-\mu}{\sigma}$.
- We get

$$\begin{aligned} E[Z] &= \int_{-\infty}^{\infty} x f_Z(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-x^2/2}dx \\ &= -\frac{1}{\sqrt{2\pi}} e^{-x^2/2}\Big|_{-\infty}^{\infty} = 0. \end{aligned}$$

- Thus,

$$\begin{aligned} \text{Var}(Z) &= E[Z^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}dx \\ &= \frac{1}{\sqrt{2\pi}}\left(-xe^{-x^2/2}\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-x^2/2}dx\right) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}dx = 1. \end{aligned}$$

- But $X = \mu + \sigma Z$. Hence $E[X] = \mu + \sigma E[Z] = \mu$ and $\text{Var}(X) = \sigma^2 \text{Var}(Z) = \sigma^2$.

Cumulative Distribution Function

- It is customary to denote the cumulative distribution function of a standard normal random variable by $\Phi(x)$:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2} dy.$$

- The values of $\Phi(x)$ for nonnegative x are given in a table.
- For negative values of x, $\Phi(x)$ can be obtained from the relationship

$$\Phi(-x) = 1 - \Phi(x), \quad -\infty < x < \infty.$$

- This equation states that if Z is a standard normal random variable, then $P\{Z \leq -x\} = P\{Z > x\}$, $-\infty < x < \infty$.
- Since $Z = \frac{X-\mu}{\sigma}$ is a standard normal random variable whenever X is normally distributed with parameters μ and σ^2, it follows that the distribution function of X can be expressed as

$$F_X(a) = P\{X \leq a\} = P\left(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Example

- If X is a normal random variable with parameters $\mu = 3$ and $\sigma^2 = 9$, find (a) $P\{2 < X < 5\}$; (b) $P\{X > 0\}$; (c) $P\{|X-3| > 6\}$.

(a)

$$\begin{aligned} P\{2 < X < 5\} &= P\{\tfrac{2-3}{3} < \tfrac{X-3}{3} < \tfrac{5-3}{3}\} \\ &= P\{-\tfrac{1}{3} < Z < \tfrac{2}{3}\} = \Phi(\tfrac{2}{3}) - \Phi(-\tfrac{1}{3}) \\ &= \Phi(\tfrac{2}{3}) - [1 - \Phi(\tfrac{1}{3})] \approx 0.3779. \end{aligned}$$

(b)

$$\begin{aligned} P\{X > 0\} &= P\{\tfrac{X-3}{3} > \tfrac{0-3}{3}\} = P\{Z > -1\} \\ &= 1 - \Phi(-1) = \Phi(1) \approx 0.8413. \end{aligned}$$

(c)

$$\begin{aligned} P\{|X-3| > 6\} &= P\{X > 9\} + P\{X < -3\} \\ &= P\{\tfrac{X-3}{3} > \tfrac{9-3}{3}\} + P\{\tfrac{X-3}{3} < \tfrac{-3-3}{3}\} \\ &= P\{Z > 2\} + P\{Z < -2\} \\ &= 1 - \Phi(2) + \Phi(-2) = 2[1 - \Phi(2)] \approx 0.0456. \end{aligned}$$

Example: Grading "On the Curve"

- An examination is frequently regarded as being good (in the sense of determining a valid grade spread for those taking it) if the test scores of those taking the examination can be approximated by a normal density function.
- In other words, a graph of the frequency of grade scores should have approximately the bell-shaped form of the normal density.
- The instructor often uses the test scores to estimate the normal parameters μ and σ^2.
- (S)he then assigns the letter grades as follows:
    - *A* to those whose test score is greater than $\mu + \sigma$;
    - *B* to those whose score is between μ and $\mu + \sigma$;
    - *C* to those whose score is between $\mu - \sigma$ and μ;
    - *D* to those whose score is between $\mu - 2\sigma$ and $\mu - \sigma$;
    - *F* to those getting a score below $\mu - 2\sigma$.
- This strategy is sometimes referred to as grading "on the curve."

Example (Cont'd)

- Note that

$$\begin{aligned} P\{X > \mu + \sigma\} &= P\{\tfrac{X-\mu}{\sigma} > 1\} = 1 - \Phi(1) \approx 0.1587; \\ P\{\mu < X < \mu + \sigma\} &= P\{0 < \tfrac{X-\mu}{\sigma} < 1\} \\ &= \Phi(1) - \Phi(0) \approx 0.3413; \\ P\{\mu - \sigma < X < \mu\} &= P\{-1 < \tfrac{X-\mu}{\sigma} < 0\} \\ &= \Phi(0) - \Phi(-1) \approx 0.3413; \\ P\{\mu - 2\sigma < X < \mu - \sigma\} &= P\{-2 < \tfrac{X-\mu}{\sigma} < -1\} \\ &= \Phi(2) - \Phi(1) \approx 0.1359; \\ P\{X < \mu - 2\sigma\} &= P\{\tfrac{X-\mu}{\sigma} < -2\} = \Phi(-2) \approx 0.0228. \end{aligned}$$

  Thus, approximately:
  - 16 percent of the class will receive an A grade;
  - 34 percent will receive a B grade;
  - 34 percent will receive a C grade;
  - 14 percent will receive a D grade;
  - 2 percent will fail.

Example

- An expert witness in a paternity suit testifies that the length (in days) of human gestation is approximately normally distributed with parameters $\mu = 270$ and $\sigma^2 = 100$.
  The defendant proves that he was out of the country between 290 days before the birth to 240 days after the birth.
  If the defendant was, in fact, the father, what is the probability that the mother could have had the very long or very short gestation indicated by the testimony?
  Let X denote the length of the gestation.
  Assume that the defendant is the father.
  The probability that the birth could occur within the indicated period is

$$\begin{aligned} P\{X > 290 \text{ or } X < 240\} &= P\{X > 290\} + P\{X < 240\} \\ &= P\{\tfrac{X-270}{10} > 2\} + P\{\tfrac{X-270}{10} < -3\} \\ &= 1 - \Phi(2) + 1 - \Phi(3) \approx 0.0241. \end{aligned}$$

Example

- Suppose that a binary message is transmitted by wire from location A to B over a wire subject to a channel noise disturbance.

  To reduce the possibility of error, the value 2 is sent when the message is 1 and the value -2 is sent when the message is 0.

  If x, $x = \pm 2$, is the value sent at A, then R, the value received at B, is given by $R = x + N$, where N is the channel noise.

  When the message is received at location B, the receiver decodes it according to the following rule:
  - If $R \geq 0.5$, then 1 is concluded.
  - If $R < 0.5$, then 0 is concluded.

  Assume the channel noise N is normally distributed.

  Determine the error probabilities (assuming N is a standard normal random variable).

Example (Cont'd)

- Two types of errors can occur:
  - The message 1 is incorrectly determined to be 0;
  - The message 0 is incorrectly determined to be 1.

  The first type of error will occur if the message is 1 and $2 + N < 0.5$.

  The second will occur if the message is 0 and $-2 + N \geq 0.5$.

  Hence,

$$\begin{aligned} P\{\text{error}|\text{message is } 1\} &= P\{N < -1.5\} \\ &= 1 - \Phi(1.5) \approx 0.0668; \\ P\{\text{error}|\text{message is } 0\} &= P\{N \geq 2.5\} \\ &= 1 - \Phi(2.5) \approx 0.0062. \end{aligned}$$

Subsection 5

Exponential Random Variables

Exponential Random Variables

- A continuous random variable whose probability density function is given, for some $\lambda > 0$, by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

  is said to be an **exponential random variable** (or, more simply, is said to be **exponentially distributed**) with parameter λ.
- We compute the cumulative distribution function $F(a)$ of an exponential random variable:

$$\begin{aligned} F(a) &= P\{X \leq a\} = \int_0^a \lambda e^{-\lambda x} dx \\ &= -e^{-\lambda x}|_0^a = 1 - e^{-\lambda a}, \quad a \geq 0. \end{aligned}$$

- This also verifies that $F(\infty) = \int_0^\infty \lambda e^{-\lambda x} dx = 1$.

Expectation and Variance of an Exponential Variable

- Let X be an exponential random variable with parameter λ.
- The density function is given by $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$
- Thus, for $n > 0$, $E[X^n] = \int_0^\infty x^n \lambda e^{-\lambda x} dx$.
- Integrating by parts (with $\lambda e^{-\lambda x} = dv$ and $u = x^n$) yields

$$\begin{aligned} E[X^n] &= -x^n e^{-\lambda x}|_0^\infty + \int_0^\infty e^{-\lambda x} n x^{n-1} dx \\ &= 0 + \frac{n}{\lambda} \int_0^\infty \lambda e^{-\lambda x} x^{n-1} dx \\ &= \frac{n}{\lambda} E[X^{n-1}]. \end{aligned}$$

- Letting $n = 1, 2$ gives $E[X] = \frac{1}{\lambda}$ and $E[X^2] = \frac{2}{\lambda} E[X] = \frac{2}{\lambda^2}$.
- Hence,

$$\text{Var}(X) = \frac{2}{\lambda^2} - (\frac{1}{\lambda})^2 = \frac{1}{\lambda^2}.$$

Example

- Suppose that the length of a phone call in minutes is an exponential random variable with parameter $\lambda = \frac{1}{10}$.

  Mr. Jack arrives just ahead of Mr. Jim at a pub's public phone.

  Find the probability that Mr. Jim will have to wait:

  (a) More than 10 minutes;
  (b) Between 10 and 20 minutes.

  Let X denote the length of the call made by Mr. Jack.

  Then the desired probabilities are

  (a) $P\{X > 10\} = 1 - F(10) = 1 - (1 - e^{-\frac{1}{10}10}) = e^{-1} \approx 0.368$.
  (b) $P\{10 < X < 20\} = F(20) - F(10) = (1 - e^{-\frac{1}{10}20}) - (1 - e^{-\frac{1}{10}10}) = e^{-1} - e^{-2} \approx 0.233$.

Memoryless Random Variables

- We say that a nonnegative random variable X is **memoryless** if

$$P\{X > s + t | X > t\} = P\{X > s\}, \text{ for all } s, t \geq 0.$$

- Thinking of X as being the lifetime of some instrument, the equation states:

  The probability that the instrument survives for at least $s + t$ hours, given that it has survived t hours, is the same as the initial probability that it survives for at least s hours.

- The equation is equivalent to

$$\frac{P\{X > s + t, X > t\}}{P\{X > t\}} = P\{X > s\}$$

$$\text{or } P\{X > s + t\} = P\{X > s\} P\{X > t\}.$$

- Exponentially distributed random variables are memoryless, since $e^{-\lambda(s+t)} = e^{-\lambda s} e^{-\lambda t}$.

Example

- Consider a post office that is staffed by two clerks.

  Suppose that when Mr. Smith arrives, he discovers that Ms. Jones is being served by one of the clerks and Mr. Brown by the other.

  Suppose also that Mr. Smith is told that his service will begin as soon as either Ms. Jones or Mr. Brown leaves.

  If the amount of time that a clerk spends with a customer is exponentially distributed with parameter λ, what is the probability that, of the three customers, Mr. Smith is the last to leave?

Example (Cont'd)

- Consider the time at which Mr. Smith first finds a free clerk.

  At this point, either Ms. Jones or Mr. Brown would have just left, and the other one would still be in service.

  However, because the exponential is memoryless, it follows that the additional amount of time that this other person (either Ms. Jones or Mr. Brown) would still have to spend in the post office is exponentially distributed with parameter λ.

  That is, it is the same as if service for that person were just starting at this point.

  Hence, by symmetry, the probability that the remaining person finishes before Smith leaves must equal $\frac{1}{2}$.

Exponential as Only Memoryless Distribution

- It turns out that not only is the exponential distribution memoryless, but it is also the unique distribution possessing this property.
- To see this, suppose that X is memoryless and let $\overline{F}(x) = P\{X > x\}$.
- Then, by the preceding equation,

$$\overline{F}(s+t) = \overline{F}(s)\overline{F}(t).$$

- That is, $\overline{F}(\cdot)$ satisfies the functional equation

$$g(s+t) = g(s)g(t).$$

- It turns out that the only right continuous solution of this functional equation is $g(x) = e^{-\lambda x}$.
- Since a distribution function is always right continuous, we must have

$$\overline{F}(x) = e^{-\lambda x} \text{ or } F(x) = P\{X \leq x\} = 1 - e^{-\lambda x}.$$

Example

- Suppose that the number of miles that a car can run before its battery wears out is exponentially distributed with an average value of 10,000 miles.

  If a person desires to take a 5000 mile trip, what is the probability that he or she will be able to complete the trip without having to replace the car battery?

  What can be said when the distribution is not exponential?

  It follows by the memoryless property of the exponential distribution that the remaining lifetime (in thousands of miles) of the battery is exponential with parameter $\lambda = \frac{1}{10}$.

  Hence, the desired probability is

$$P\{\text{remaining lifetime} > 5\} = 1 - F(5) = e^{-5\lambda} = e^{-1/2} \approx 0.604.$$

Example (Cont'd)

- Suppose that the lifetime distribution F is not exponential.

  Let t be the number of miles that the battery had been in use prior to the start of the trip.

  Then the relevant probability is

$$P\{\text{lifetime} > t+5 | \text{lifetime} > t\} = \frac{1-F(t+5)}{1-F(t)}.$$

- Therefore, if the distribution is not exponential, additional information is needed (namely, the value of t) before the desired probability can be calculated.

Laplace Distribution

- A variant of the exponential is the distribution of a random variable that is equally likely to be either positive or negative and whose absolute value is exponentially distributed with parameter λ, $\lambda \geq 0$.
- Such a random variable is said to have a **Laplace distribution**, and its density is given by

$$f(x) = \frac{1}{2}\lambda e^{-\lambda|x|}, \quad -\infty < x < \infty.$$

- Its distribution function is given by

$$\begin{aligned} F(x) &= \begin{cases} \frac{1}{2}\int_{-\infty}^{x} \lambda e^{\lambda x} dx, & x < 0 \\ \frac{1}{2}\int_{-\infty}^{0} \lambda e^{\lambda x} dx + \frac{1}{2}\int_{0}^{x} \lambda e^{-\lambda x} dx, & x > 0 \end{cases} \\ &= \begin{cases} \frac{1}{2}e^{\lambda x}, & x < 0 \\ 1 - \frac{1}{2}e^{-\lambda x}, & x > 0 \end{cases} \end{aligned}$$

Example

- Suppose that a binary message is to be transmitted from A to B, with the value 2 being sent when the message is 1 and -2 when it is 0.

  However, suppose now that, rather than being a standard normal random variable, the channel noise N is a Laplacian random variable with parameter $\lambda = 1$.

  Suppose again that if R is the value received at location B, then the message is decoded as follows:
  - If $R \geq 0.5$, then 1 is concluded.
  - If $R < 0.5$, then 0 is concluded.

  In this case, where the noise is Laplacian with parameter $\lambda = 1$, the two types of errors will have probabilities given by:

$$P\{\text{error}|\text{message 1 is sent}\} = P\{N < -1.5\} = \frac{1}{2}e^{-1.5} \approx 0.1116;$$
$$P\{\text{error}|\text{message 0 is sent}\} = P\{N \geq 2.5\} = \frac{1}{2}e^{-2.5} \approx 0.041.$$

Hazard Rate Functions

- Consider a positive continuous random variable X that we interpret as being the lifetime of some item.
- Let X have distribution function F and density f.
- The **hazard rate** (sometimes called the **failure rate**) **function** $\lambda(t)$ of F is defined by

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)}, \quad \text{where } \overline{F} = 1 - F.$$

- To interpret $\lambda(t)$, suppose that the item has survived for a time t and we desire the probability that it will not survive for an additional time dt. That is, consider $P\{X \in (t, t+dt) | X > t\}$.

$$\begin{aligned} P\{X \in (t, t+dt) | X > t\} &= \frac{P\{X \in (t,t+dt), X > t\}}{P\{X > t\}} \\ &= \frac{P\{X \in (t,t+dt)\}}{P\{X > t\}} \approx \frac{f(t)}{\overline{F}(t)} dt. \end{aligned}$$

- Thus, $\lambda(t)$ represents the conditional probability intensity that a t-unit-old item will fail.

Rate of the Distribution

- Suppose now that the lifetime distribution is exponential.
- By the memoryless property, the distribution of remaining life for a t-year-old item is the same as that for a new item.
- Hence, $\lambda(t)$ should be constant.
- In fact, we get
$$\lambda(t) = \frac{f(t)}{\overline{F}(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda.$$
- Thus, the failure rate function for the exponential distribution is constant.
- The parameter λ is often referred to as the **rate** of the distribution.

Distribution from Failure Rate

- It turns out that the failure rate function $\lambda(t)$ uniquely determines the distribution F.
- To prove this, note that, by definition, $\lambda(t) = \frac{f(t)}{\overline{F}(t)} = \frac{\frac{d}{dt}F(t)}{1-F(t)}$.
- Integrating both sides yields

$$\log(1 - F(t)) = -\int_0^t \lambda(t)dt + k$$
$$1 - F(t) = e^k \exp\{-\int_0^t \lambda(t)dt\}.$$

- Letting $t = 0$ shows that $k = 0$.
- Thus,

$$F(t) = 1 - \exp\{-\int_0^t \lambda(t)dt\}.$$

Linear Hazard Rate Function

- Suppose a random variable has a linear hazard rate function, i.e.,

$$\lambda(t) = a + bt.$$

- Then its distribution function is given by

$$F(t) = 1 - \exp\{-\int_0^t \lambda(t)dt\} = 1 - e^{-at-bt^2/2}.$$

- Differentiation yields its density, namely,

$$f(t) = (a+bt)e^{-(at+bt^2/2)}, \quad t \geq 0.$$

- When $a = 0$, the preceding equation is known as the **Rayleigh density function**.

Example

- One often hears that the death rate of a person who smokes is, at each age, twice that of a nonsmoker.

  Does that mean that a nonsmoker has twice the probability of surviving a given number of years as does a smoker of the same age?

  Suppose that:
  - $\lambda_s(t)$ denotes the hazard rate of a smoker of age t;
  - $\lambda_n(t)$ the hazard rate of a nonsmoker of age t.

  The statement is equivalent to the statement $\lambda_s(t) = 2\lambda_n(t)$.

  The probability that an A-year-old nonsmoker will survive until age B, $A < B$, is

$$\begin{aligned} & P\{A\text{-year-old nonsmoker reaches age } B\} \\ &= P\{\text{nonsmoker's lifetime} > B | \text{nonsmoker's lifetime} > A\} \\ &= \frac{1-F_{\text{non}}(B)}{1-F_{\text{non}}(A)} = \frac{\exp\{-\int_0^B \lambda_n(t)dt\}}{\exp\{-\int_0^A \lambda_n(t)dt\}} = \exp\{-\int_A^B \lambda_n(t)dt\}. \end{aligned}$$

Example (Cont'd)

- The corresponding probability for a smoker is, by the same reasoning,

$$\begin{aligned} & P\{A\text{-year-old smoker reaches age } B\} \\ & = \exp\left\{-\int_A^B \lambda_s(t)dt\right\} = \exp\left\{-2\int_A^B \lambda_n(t)dt\right\} \\ & = \left[\exp\left\{-\int_A^B \lambda_n(t)dt\right\}\right]^2. \end{aligned}$$

- In other words, for two people of the same age, one of whom is a smoker and the other a nonsmoker, the probability that the smoker survives to any given age is the square (not one-half) of the corresponding probability for a nonsmoker.

Subsection 6

The Distribution of a Function of a Random Variable

Example

- Let X be uniformly distributed over $(0,1)$.
- We obtain the distribution of the random variable Y, defined by $Y = X^n$, as follows:
- For $0 \leq y \leq 1$,

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} = P\{X^n \leq y\} \\ &= P\{X \leq y^{1/n}\} = F_X(y^{1/n}) = y^{1/n}. \end{aligned}$$

- For instance, the density function of Y is given by

$$f_Y(y) = \begin{cases} \frac{1}{n} y^{1/n-1}, & 0 \leq y \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Example

- Let X be a continuous random variable with probability density f_X.
- The distribution of $Y = X^2$ is obtained as follows:
- For $y \geq 0$,

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} \\ &= P\{X^2 \leq y\} \\ &= P\{-\sqrt{y} \leq X \leq \sqrt{y}\} \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

- Differentiation yields

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})].$$

Example

- Let X have a probability density f_X.
- Then $Y = |X|$ has a density function that is obtained as follows:
- For $y \geq 0$,

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} \\ &= P\{|X| \leq y\} \\ &= P\{-y \leq X \leq y\} \\ &= F_X(y) - F_X(-y). \end{aligned}$$

- Hence, on differentiation, we obtain

$$f_Y(y) = f_X(y) + f_X(-y), \quad y \geq 0.$$

Probability Density of a Function of a Variable

Theorem

Let X be a continuous random variable having probability density function f_X. Suppose that $g(x)$ is a strictly monotonic (increasing or decreasing), differentiable (and thus continuous) function of x. Then the random variable Y defined by $Y = g(X)$ has a probability density function:

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)]\left|\frac{d}{dy}g^{-1}(y)\right|, & \text{if } y = g(x) \text{ for some } x \\ 0, & \text{if } y \neq g(x) \text{ for all } x \end{cases}$$

where $g^{-1}(y)$ is defined to equal that value of x such that $g(x) = y$.

- We prove the theorem when $g(x)$ is an increasing function.
  Suppose that $y = g(x)$ for some x.
  Then, with $Y = g(X)$,

$$F_Y(y) = P\{g(X) \leq y\} = P\{X \leq g^{-1}(y)\} = F_X(g^{-1}(y)).$$

Probability Density of a Function of a Variable (Cont'd)

- We found $F_Y(y) = F_X(g^{-1}(y))$.
  Differentiation gives
  $$f_Y(y) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y).$$
  Under the hypothesis that $g^{-1}(y)$ is nondecreasing, $\frac{d}{dy} g^{-1}(y) \geq 0$.
  Hence, the formula agrees with the one in the statement.
  When $y \neq g(x)$ for any x, then $F_Y(y)$ is either 0 or 1.
  In either case $f_Y(y) = 0$.

Example

- Suppose
  - X is a continuous nonnegative random variable with density f;
  - $Y = X^n$.

Find f_Y, the probability density function of Y.

If $g(x) = x^n$, then

$$g^{-1}(y) = y^{1/n}$$

and

$$\frac{d}{dy} g^{-1}(y) = \frac{1}{n} y^{1/n-1}.$$

Hence, from the theorem, we obtain $f_Y(y) = \frac{1}{n} y^{1/n-1} f(y^{1/n})$.

For $n = 2$, this gives

$$f_Y(y) = \frac{1}{2\sqrt{y}} f(\sqrt{y}).$$

Since $X \geq 0$, this agrees with a previous example.