STEVENS
INSTITUTE OF TECHNOLOGY
1870

MA 540: Introduction to Probability Theory

Lecture 5: Continuous random variables

Benjamin Leinwand

Class business

- **There is a poll due Saturday to give feedback about the class**
- **HW 4 is due Thursday October 12th instead of October 11th**
- **HW 5 is due Thursday October 19th instead of October 18th**
- **Midterm is October 23rd**
- The end of the October 16th class will be reserved for student led midterm review Q&A
  - I will make old midterms available on canvas for your review

Exam info

- October 23rd during class
- Multiple choice, some proofs (like from class, but not the very long ones), some computational problems as in the homeworks
  - See the previous midterms that will be on Canvas for an idea of what to expect
- A single 2 sided standard size handwritten sheet of notes is allowed, no other aids
  - Anything printed, even if it was originally handwritten, is forbidden
- Will cover everything from the beginning of the semester until the end of Lecture 5 notes

Continuous RV

- An RV which takes on an uncountable number of possible values
- Nonnegative function f for all real x in (–inf, inf) such that for any set B of real numbers

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

Where f is called the probability density function of the RV X

The pdf corresponds to the pmf for discrete RVs, but as there are possibly uncountable values with positive probability, we cannot take the value of the pdf itself, but instead need this integral

Probability Density Function and CDF for continuous RV

- To be a valid probability the pdf must satisfy

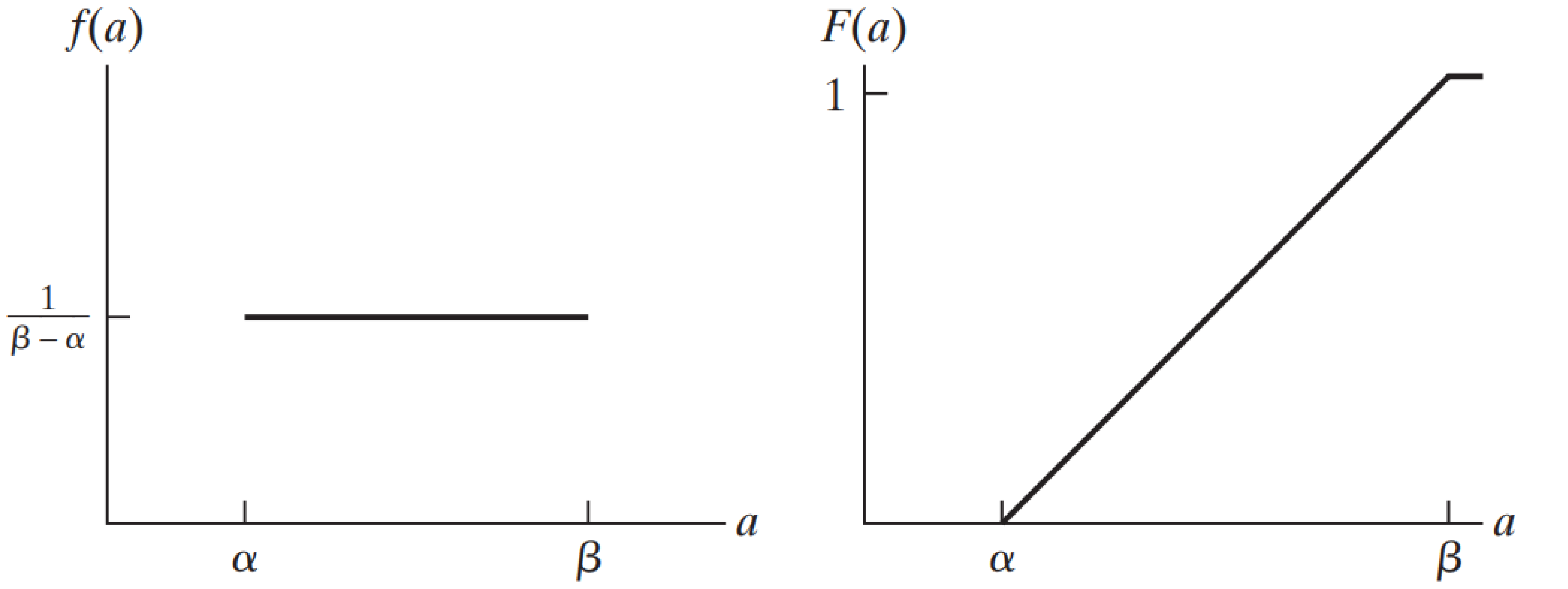
$$1 = P\{X \in (-\infty, \infty)\} = \int_{-\infty}^{\infty} f(x)\, dx$$

- If B = [a, b], $P\{a \le X \le b\} = \int_a^b f(x)\, dx$
  - If a=b, $P\{X = a\} = \int_a^a f(x)\, dx = 0$ which also explains why the term above can include two less than or equal symbols (contrast with PMF)

- Corollary: CDF of continuous RV is $P\{X < a\} = P\{X \le a\} = F(a) = \int_{-\infty}^{a} f(x)\, dx$

Graphs of pdf and CDF of Uniform Distribution

- For a uniform random variable on the interval (α, β), we show the pdf f(a), and CDF F(a)



Continuous RV Example 1

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 $P\{X > 1\}$.

Continuous RV Example 1

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 $P\{X > 1\}$.

Solution (a) Since f is a probability density function, we must have $\int_{-\infty}^{\infty} f(x)\,dx = 1$, implying that

$$C\int_0^2 (4x - 2x^2)\,dx = 1$$

or

$$C\left[2x^2 - \frac{2x^3}{3}\right]\Big|_{x=0}^{x=2} = 1$$

or

$$C = \frac{3}{8}$$

Hence,

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

Continuous RV Example 2

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

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

What is 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?

Continuous RV Example 1

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

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

What is 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?

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \lambda \int_{0}^{\infty} e^{-x/100}\,dx = -\lambda(100)e^{-x/100}\Big|_0^{\infty} = 100\lambda \quad \text{or} \quad \lambda = \frac{1}{100}$$

$$P\{50 < X < 150\} = \int_{50}^{150} \frac{1}{100} e^{-x/100}\,dx = -e^{-x/100}\Big|_{50}^{150}$$
$$= e^{-1/2} - e^{-3/2} \approx .383$$

$$P\{X < 100\} = \int_{0}^{100} \frac{1}{100} e^{-x/100}\,dx = -e^{-x/100}\Big|_0^{100} = 1 - e^{-1} \approx .632$$

Getting the density of an RV that is a function of another RV

- What if we have the pdf of X, but want the pdf of Y = g(X)?
  - Operate using the CDF (not the PDF!), then differentiate.

Example: If X is continuous with distribution function F_X and density function f_X, find the density function of Y = 2X

$$\begin{aligned} F_Y(a) &= P\{Y \leq a\} \\ &= P\{2X \leq a\} \end{aligned}$$

Getting the density of an RV that is a function of another RV

- What if we have the pdf of X, but want the pdf of Y = g(X)?
  - **Operate using the CDF (not the PDF!), then differentiate**

Example: If X is continuous with distribution function F_X and density function f_X, find the density function of Y = 2X

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

Recall: $f(a) = \frac{d}{da}F(a)$
So take the derivative with respect to a

Note: this function is not obvious using the pdf alone. We have to use the CDF, and a bit of logic

$$f_Y(a) = \frac{1}{2} f_X(a/2)$$

Expectation for Continuous RVs

- Recall for a discrete RV $E[X] = \sum_x xP\{X = x\}$
- The analogous definition for a continuous RV is $E[X] = \int_{-\infty}^{\infty} xf(x)\,dx$
- Example: Find $E[X]$ when the density function of X is

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

$$E[X] = \int xf(x)\,dx \quad = \int_0^1 2x^2\,dx \quad = \frac{2}{3}$$

Expected Value of Function of RV Example

The density function of X is given by

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

Find $E[e^X]$.

Again, we first use the CDF, logic, and rules of calculus to get the pdf. Then we get the expectation based on the pdf we recover

Let $Y = e^X \quad \text{for } 1 \le x \le e,$

$$\begin{aligned} F_Y(x) &= P\{Y \le x\} \\ &= P\{e^X \le x\} \\ &= P\{X \le \log(x)\} \\ &= \int_0^{\log(x)} f(y)\,dy \\ &= \log(x) \end{aligned}$$

$$f_Y(x) = \frac{1}{x} \qquad 1 \le x \le e$$

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

$$\begin{aligned} E[e^X] = E[Y] &= \int_{-\infty}^{\infty} x f_Y(x)\,dx \\ &= \int_1^e dx \\ &= e - 1 \end{aligned}$$

Proposition with application to previous problem: Expected Value of Function of RV

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

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

- In the previous example, that gives

$$\begin{aligned} E[e^X] &= \int_0^1 e^x\,dx \qquad \text{since } f(x) = 1, \quad 0 < x < 1 \\ &= e - 1 \end{aligned}$$

- Note: we will only be proving this for a nonnegative RV g(X), but it is true in general

Lemma leading to limited proposition

- For a nonnegative random variable Y, $E[Y] = \int_0^\infty P\{Y > y\}\, dy$
- **Proof** $\int_0^\infty P\{Y > y\}\, dy = \int_0^\infty \int_y^\infty f_Y(x)\, dx\, dy$

We can interchange the order of integration because the function is non-negative (Tonelli's theorem)

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

Proof of proposition

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

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

- From the previous lemma, for any nonnegative function g

$$E[g(X)] = \int_0^{\infty} P\{g(X) > y\}\,dy = \int_0^{\infty} \int_{x:g(x)>y} f(x)\,dx\,dy$$

$$= \int_{x:g(x)>0} \int_0^{g(x)} dy\, f(x)\,dx = \int_{x:g(x)>0} g(x)f(x)\,dx \qquad \blacksquare$$

Corollary and variance

- If a and b are constants: $E[aX + b] = aE[X] + b$
  - Same proof as for discrete case, just replace summations with integrals

- As with a discrete RV, the variance for a continuous RV is defined as

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

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

  - Showing these are equivalent is again the same as in the discrete case, just replace summations with integrals

Calculating Variance Example

- Find Var(X) for X with the following pdf

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

- Hint: We showed earlier that for this RV, E[X] = 2/3

Calculating Variance Example

- Find Var(X) for X with the following pdf

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

- We showed earlier that E[X] = 2/3
- First way $E[X - E[X]]^2 = \int_0^1 (x - E[X])^2 f(x)dx = \int_0^1 (x - \frac{2}{3})^2\, 2x\, dx = 2\int_0^1 x^3 - \frac{4}{3}x^2 + \frac{4}{9}x\, dx$

$$= 2[\frac{1}{4}x^4 - \frac{4}{9}x^3 + \frac{4}{18}x^2 \;|_0^1\;] = 2[\frac{9}{36} - \frac{16}{36} + \frac{8}{36}] = \frac{1}{18}$$

Second way: $E[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 2x^3 dx = \frac{1}{2}x^4 \;|_0^1 = ½$

So $E[X^2] - [E[X]]^2 = 1/2 - (2/3)^2 = 1/18$

Uniform RV

- A random variable which is uniformly distributed over the interval (0, 1) if its probability density function is given by

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

- f(x) is nonnegative and $\int_{-\infty}^{\infty} f(x)dx = \int_0^1 dx = 1$
- Note for any 0 < a < b < 1

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

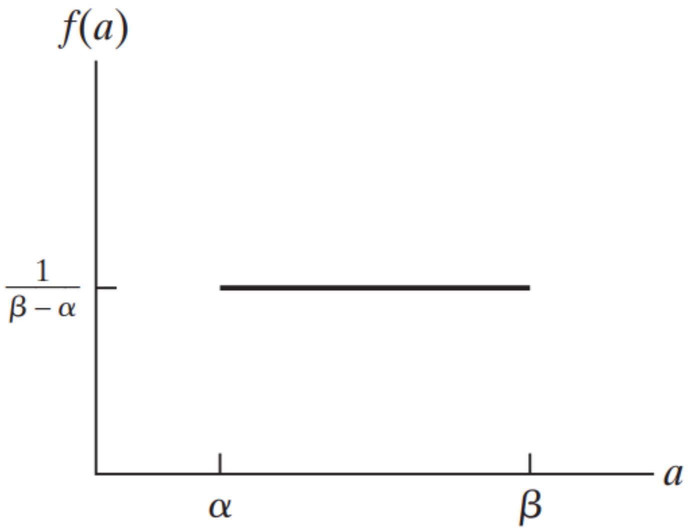
- **General uniform distribution**: X is a uniform random variable on the interval (α, β) if

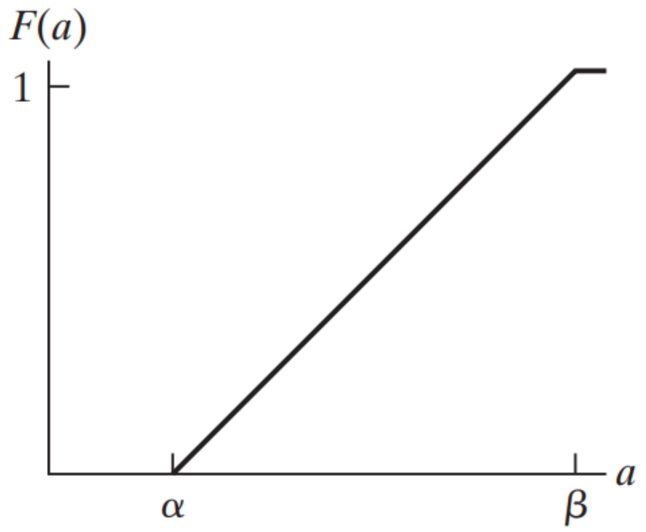
$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{if } \alpha < x < \beta \\ 0 & \text{otherwise} \end{cases}$$

CDF for uniform and graph of pdf and CDF

- For a general uniform random variable on the interval (α, β), the CDF is given by

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





Expectation and variance for uniform

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f(x)\, dx \\
&= \int_{\alpha}^{\beta} \frac{x}{\beta - \alpha}\, dx \\
&= \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)}
\end{aligned}
$$

$$
\boxed{= \frac{\beta + \alpha}{2}}
$$

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

$$
\mathrm{Var}(X)
$$

$$
= \frac{\beta^2 + \alpha\beta + \alpha^2}{3} - \frac{(\alpha + \beta)^2}{4}
$$

$$
\boxed{= \frac{(\beta - \alpha)^2}{12}}
$$

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

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
- a) describes one of two situations: arriving between 7:10 and 7:15, or between 7:25 and 7:30

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

- b) describes arriving between 7:00 and 7:05, or between 7:15 and 7:20

$$P\{0 < X < 5\} + P\{15 < X < 20\} = \frac{1}{3}$$

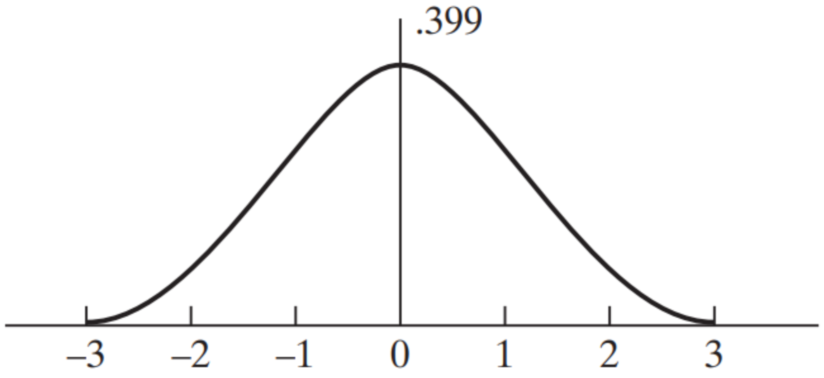
Normal RVs

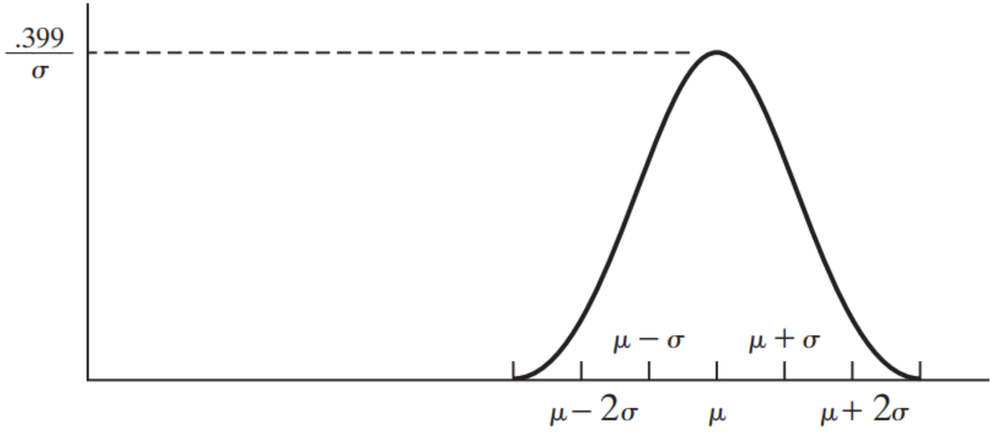
- Normal distributions have 2 parameters, μ and σ^2 and the density of any normal random variable is

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

- $\frac{1}{\sqrt{2\pi}\sigma}$ can be seen as a normalizing constant to ensure this is a valid distribution, while

$e^{-(x-\mu)^2/2\sigma^2}$ mandates that the farther you get from μ in either direction, the smaller f(x) gets.





Normal RV is a probability

- Must show $\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2}\,dx = 1$ Substitute y = (x − μ)/σ

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

suffices to show $\int_{-\infty}^{\infty} e^{-y^2/2}\,dy = \sqrt{2\pi}$

- Let $I = \int_{-\infty}^{\infty} e^{-y^2/2}\,dy.$

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

- Convert to polar coordinates: x = r cos θ, y = rsin θ, and dy dx = r dθ dr

$$I^2 = \int_0^{\infty}\int_0^{2\pi} e^{-r^2/2} r\,d\theta \quad = 2\pi \int_0^{\infty} re^{-r^2/2} dr = -2\pi e^{-r^2/2}\Big|_0^{\infty} \quad = 2\pi$$

- As $I^2 = 2\pi$, $I = \sqrt{2\pi}$ so a Normal RV is a probability

Linear transformation of Normal

- If X is a normal random variable with parameters μ and σ^2 then if we take Y = aX + b, then Y is also a normal distribution with parameters aμ + b and $a^2\sigma^2$
- Proof:

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

Again, start with CDF, change variables, differentiate to get to PDF

Differentiate to get f(y)

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

- Upshot: it's easy to convert any normal CDF into a N(0,1) CDF and vice versa. To turn arbitrary Normal CDF X into a standard normal (with μ= 0 and σ = 1), let Y = (X − μ)/σ

Expectation of Normal

- Let $Z = (X - \mu)/\sigma$

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

- $X = \sigma Z + \mu$ so using the rule about expectations of a function of an RV

$$E[X] = E[\sigma Z + \mu] = \sigma E[Z] + \mu = \mu$$

Variance of normal

- Again, letting Z = (X – μ)/σ, since E[Z] = 0, Var[Z] = E[Z^2] = $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\, dx$

- Use integration by parts with u = x and dv = $xe^{-x^2/2}$, du = dx and v = - $e^{-x^2/2}$

$$= \frac{1}{\sqrt{2\pi}} \left(-xe^{-x^2/2} |_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-x^2/2}\, dx \right)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2}\, dx$$

$$= 1$$

We derived the value of this integral before when showing a Normal is a probability distribution. There we called it *I* and showed it equals $\sqrt{2\pi}$ by squaring it and using polar coordinates

- X = σZ + μ so using a rule about variances,

Var(X) = σ^2Var(Z) = σ^2

CDF of Normal and a table and symmetry

- The symbol for the CDF of a standard Normal is $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy$
  - The CDF of a standard normal is often kept in a Z-table
- Looking at the formula $\mathrm{f(x)} = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-(x^2)}{2}\right)$, it's easy to see that for a standard Normal, f(x) = f(-x). This implies that the standard Normal is symmetric about 0. It also implies that Φ(x) = 1-Φ(-x)
  - In other words, if Z is a standard normal RV, P(Z< -z) = P(Z > z)
- Finally, for a general Normal RV with parameters μ and σ^2:

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

  - A Z-table can be referenced after converting a value from an arbitrary normal to a Z-score

Normal Probability Example 1

- 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) $P\{|X - 3| > 6\}$

Normal Probability Example 1

- 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) is $P\{-1/3 < Z < 2/3\} = .3779$

(b) is $P\{Z > -1\} = .8413$ Can also do $1-P\{Z<-1\}$ or $P\{Z<1\}$

(c) is $P\{Z<-2\} + P\{Z>2\} = 2P\{Z>2\} = .0456$

To get the probability values, consult a Z-table

Normal Probability Example 2

- 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 in the suit is able to prove that he was out of the country during a period that began 290 days before the birth of the child and ended 240 days before the birth. If the defendant was, in fact, the father of the child, what is the probability that the mother could have had the very long or very short gestation indicated by the testimony?

Normal Probability Example 2

- An expert witness in a paternity suit testifies that the length (in days) of human gestation is approximately normally distributed with parameters μ = 270 and $\sigma^2 = 100$. The defendant in the suit is able to prove that he was out of the country during a period that began 290 days before the birth of the child and ended 240 days before the birth. If the defendant was, in fact, the father of the child, what is the probability that the mother could have had the very long or very short gestation indicated by the testimony?

- P{X > 290 or X < 240} = P{X > 290} + P{X < 240} = $P\left\{Z > \frac{290-270}{10}\right\} + P\left\{Z < \frac{240-270}{10}\right\}$

= $P\{Z > 2\} + P\{Z < -3\}$ = .0241

Normal Probability Example 3

- Suppose that a binary message—either 0 or 1—must be transmitted by wire from location A to location B. However, the data sent over the wire are subject to a channel noise disturbance, so, to reduce the possibility of error, the value 2 is sent over the wire 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 location A, then R, the value received at location B, is given by $R = x + N$, where N is the channel noise disturbance. When the message is received at location B, the receiver decodes it according to the following rule: If $R \geq .5$, then 1 is concluded. If $R < .5$, then 0 is concluded. Because the channel noise is often normally distributed, we will determine the error probabilities when N is a standard normal random variable

- 2 possible errors. 1) The message is 1 and we conclude 0. 2) The message is 2 and we conclude 1.
  - P(conclude 0 | message is 1} = $P(2+ N < .5\} = P\{N<-1.5\} = \Phi(-1.5) = .0668$
  - P(conclude 1 | message is 0} = $P(-2+ N \geq .5\} = P\{N \geq -2.5\} = 1- \Phi(2.5) = .0062$

The DeMoivre–Laplace limit theorem

- If Sn denotes the number of successes that occur when n independent trials, each resulting in a success with probability p, are performed, then, for any a < b,

$$P\left\{a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right\} \to \Phi(b) - \Phi(a) \quad \text{as } n \to \infty$$

- Take the empirical value, subtract the theoretical mean, and divide by the theoretical standard deviation
  - In practice we do not know the true p, so we know neither the expectation nor the standard deviation
- What does this mean? It means that as the number of trials goes to infinity, a binomial distribution can be mapped to a standard normal distribution
- More general case is Central Limit Theorem

Example using the limit theorem

- The ideal size of a first-year class at a particular college is 150 students. The college, knowing from past experience that, on the average, only 30 percent of those accepted for admission will actually attend, uses a policy of approving the applications of 450 students. Compute the probability that more than 150 first-year students attend this college.
- Note: we are using a normal distribution (which is continuous) to approximate a binomial (which is discrete), so we should use a correction, instead of writing $P\{X = i\}$ we should use $P\{i- .5 < X < i+.5\}$
  - This is called a continuity correction

Example using the limit theorem

- The ideal size of a first-year class at a particular college is 150 students. The college, knowing from past experience that, on the average, only 30 percent of those accepted for admission will actually attend, uses a policy of approving the applications of 450 students. Compute the probability that more than 150 first-year students attend this college.
- Note: we are using a normal distribution (which is continuous) to approximate a binomial (which is discrete), so we should use a correction, instead of writing $P\{X = i\}$ we should use $P\{i- .5 < X < i+.5\}$
  - This is called a continuity correction
- n=450, p=.3. Instead of $P\{X\geq 150\}$, use $P\{X\geq 150.5\}$, and use the previous formula

$$
\begin{aligned}
P\{X \geq 150.5\} &= P\left\{\frac{X - (450)(.3)}{\sqrt{450(.3)(.7)}} \geq \frac{150.5 - (450)(.3)}{\sqrt{450(.3)(.7)}}\right\} \\
&\approx 1 - \Phi(1.59) \\
&\approx .0559
\end{aligned}
$$

Exponential Random Variables

- A continuous random variable is exponentially distributed if its probability density function is given, for some λ > 0, by

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

- Uses: time until some specific event occurs – next customer, bus, etc
- CDF – Looks like what we had last week in calculating time to Poisson event – e.g. earthquake

$$1 - e^{-\lambda a} \qquad a \geq 0$$

Why?

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

Expectation and variance of an exponential RV

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

so

$$E[X^n] = \int_0^\infty x^n \lambda e^{-\lambda x}\, dx$$

for n>0

Integration by parts with $\lambda e^{-\lambda x}$ = dv and u = x^n

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

Plugging in n= 1 and 2 and solving gives

$$\boxed{E[X] = \frac{1}{\lambda}}$$

$$E[X^2] = \frac{2}{\lambda}E[X] = \frac{2}{\lambda^2}$$

$$\boxed{\mathrm{Var}(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}}$$

Exponential Example

- Suppose that the length of a phone call in minutes is an exponential random variable with parameter $\lambda = 1/10$. If someone arrives immediately ahead of you at a public telephone booth, find the probability that you will have to wait
  - (a) more than 10 minutes;
  - (b) between 10 and 20 minutes

Exponential Example

- Suppose that the length of a phone call in minutes is an exponential random variable with parameter $\lambda = 1/10$. If someone arrives immediately ahead of you at a public telephone booth, find the probability that you will have to wait
  - (a) more than 10 minutes;
  - (b) between 10 and 20 minutes

$$\begin{aligned} P\{X > 10\} &= 1 - F(10) \\ &= e^{-1} \approx .368 \end{aligned}$$

$$\begin{aligned} P\{10 < X < 20\} &= F(20) - F(10) \\ &= e^{-1} - e^{-2} \approx .233 \end{aligned}$$

Memorylessness

- We say that a nonnegative random variable X is memoryless if
  P{X > s + t | X > t} = P{X > s} for all s, t ≥ 0

- In other words $\frac{P\{X > s + t, X > t\}}{P\{X > t\}} = P\{X > s\}$ or $P\{X > s + t\} = P\{X > s\}P\{X > t\}$

- For exponential RV with arbitrary λ, this clearly holds because exp(-λ(s+t)) = exp(-λ(s)) * exp(-λ(t))

- Upshot: If the time between trains is distributed as an exponential random variable and you expect to wait 10 minutes for a train to arrive, once you've waited 10 minutes (or 20 minutes, or 500 minutes!), you expect to wait another 10 minutes for the train to arrive

Example

- Consider a post office that is staffed by two clerks. Suppose that when Mr. Smith enters the system, 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 the post office?

Example

- Consider a post office that is staffed by two clerks. Suppose that when Mr. Smith enters the system, 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 the post office?

- Memorylessness!
- At time T_1 when the first of Ms. Jones or Mr. Brown leaves. At that point, Mr. Smith can start. But because the exponential is memoryless, the time remaining for Mr. Smith and whoever is left are both exponentially distributed with parameter λ, so by symmetry, there is a ½ chance Mr. Smith is last to leave

Another Exponential 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?

$$P\{\text{remaining lifetime} > 5\} = 1 - F(5) = e^{-5\lambda} = e^{-1/2} \approx .607$$

Laplace distribution (extend the exponential to the whole real line)

- The exponential distribution only takes on nonnegative values, but what if we'd like to extend the idea to values that can be either positive or negative?
  - In this case, the absolute value of the random variable would be exponentially distributed with parameter λ

- Laplace distribution has density

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

  - The distribution is symmetric about 0, with the positive side having f(x) being half as large as the corresponding value as the exponential distribution

Gamma distribution

- A random variable is said to have a gamma distribution with parameters (α, λ), λ > 0, α > 0, if its density function is given by

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

  - Where $\Gamma(\alpha) = \int_0^{\infty} e^{-y} y^{\alpha-1}\, dy$ is called the "gamma function"
  - The Gamma function is a generalization of the factorial function to non-integer numbers

- Uses: amount of time one has to wait until a total of a events has occurred (if a is an integer)

- E[X] = a/ λ Var(X) = a / λ^2

Relationships to other distributions

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- When a=1, the gamma RV is an exponential RV
  - You can view the exponential RV as a continuous form of the geometric distribution, and the gamma as a continuous version of the negative binomial
- The Erlang distribution arises when a is a positive integer
- When $\lambda = ½$ and a = n/2 where n is a positive integer, this is a chi-squared distribution with n degrees of freedom, denoted χ_n^2
- If X ~ Gamma (a, θ) and Y ~ Gamma(b, θ) (this uses an alternative parametrization of Gamma), the X/(X+Y) ~ Beta(a, b)

Beta distribution

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Where $$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$$

- Can model a random phenomenon whose set of possible values is [0, 1]
  - For this purpose, it is most often known for being the conjugate prior for a binomial distribution
- When a = b, the beta is symmetric about 1/2, with more weight near the center as a=b increases

$$E[X] = \frac{a}{a+b}$$

$$\mathrm{Var}(X) = \frac{ab}{(a+b)^2(a+b+1)}$$

- Upshot: as a and b increase, the distribution becomes more peaked, that is, the variance shrinks

Beta distribution

$$f(x) = \begin{cases} \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1} & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Where $$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx$$

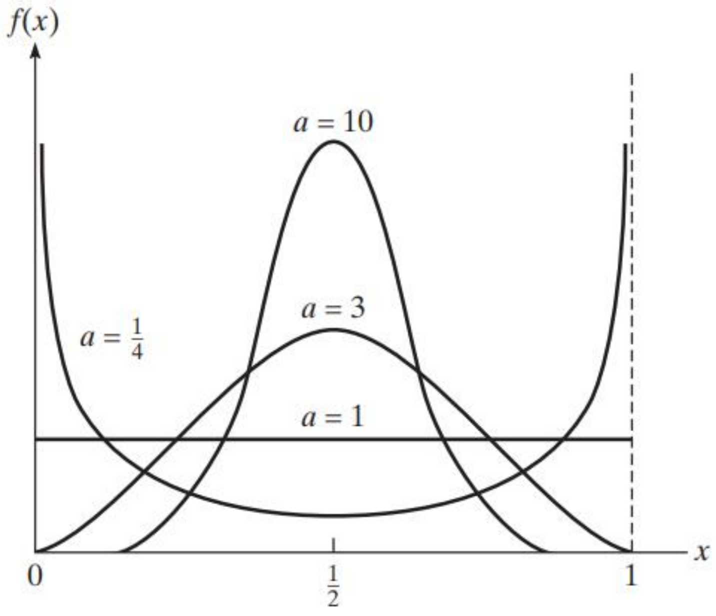


Figure 5.8 Beta densities with parameters (a, b) when $a = b$.

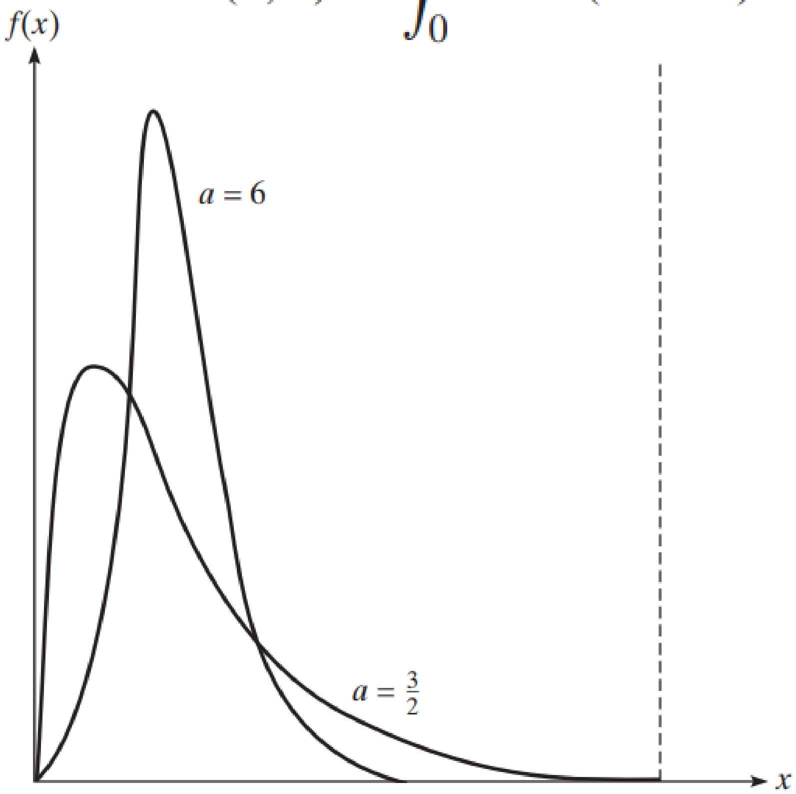


Figure 5.9 Beta densities with parameters (a, b) when $a/(a + b) = 1/20$.

Distribution of function of RV Example

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

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

Differentiate to get

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

Theorem for density of function of RV

We previously showed how to calculate the expectation of the function of an RV. Now we calculate the whole density (that is, the pdf at every point)

- 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 given by

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

where $g^{-1}(y)$ is defined as the value of x such that g(x) = y

Proof: let y = g(x) for some x, so with Y = g(X)

$$\begin{aligned} F_Y(y) &= P\{g(X) \leq y\} \\ &= P\{X \leq g^{-1}(y)\} \\ &= F_X(g^{-1}(y)) \end{aligned}$$

Differentiate to get $f_Y(y) = f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y)$

Calculating the density of a function of an RV Example

- Let X be a continuous nonnegative random variable with density function f, and let $Y = X^n$. Find f_Y, the probability density function of Y

- **Solution:** if $g(x) = x^n$ then

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

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

For $y \geq 0$, $f_Y(y) = f_X[g^{-1}(y)]\left|\frac{d}{dy}g^{-1}(y)\right|$ $\qquad f_Y(y) = \frac{1}{n}y^{1/n-1}f(y^{1/n})$

- Note for n = 2, we get $f_Y(y) = \frac{1}{2\sqrt{y}}f(\sqrt{y})$
  - As n is non-negative, this is the same value as we got for the example 2 slides ago

Cauchy distribution

$$f(x) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2} \qquad -\infty < x < \infty$$

- Though the Cauchy Distribution looks in some sense bell shaped like the Normal, it has heavier tails, and the Cauchy distribution has undefined mean and variance

- If U and V are i.i.d. N(0,1) RVs, U/V has a Cauchy distribution

- The sample mean of several i.i.d. Cauchy RVs has the same distribution as a single one of these RVs
  - Consequently, estimation does not improve with more samples

Homework 5

- Poll 2 Due Saturday 2/18

- HW 5 Due Friday 3/3
- Problems from textbook Chapter 5:
  - 5.7
  - 5.13
  - 5.16
  - 5.21
  - 5.32
  - 5.40