MATH 3339
Statistics for the Sciences
Live Lecture Help

James West
jdwest@uh.edu

University of Houston

Session 8

Office Hours: see schedule in the "Office Hours" channel on Teams
Course webpage: www.casa.uh.edu

Email policy

When you email me you **MUST** include the following

- MATH 3339 Section 20024 and a description of your issue in the **Subject Line**
- Your name and ID# in the **Body**
- Complete sentences, punctuation, and paragraph breaks
- Email messages to the class will be sent to your Exchange account (user@cougarnet.uh.edu)

Using R and R-Studio

1. Download R from https://cran.r-project.org/
2. Download R-Studio from https://www.rstudio.com/

Outline

Updates and Announcements

- Test 1 grading is ongoing
- Once visible add Test 1 to Test 1 FR for your total.

Definition of a Density Function

- A **density function** is a nonnegative function f defined of the set of real numbers such that:
$$\int_{-\infty}^{\infty} f(x)dx = 1. \qquad f(x) \neq P(X=x)$$

- If f is a density function, then its integral $F(x) = \int_{-\infty}^{x} f(u)du$ is a continuous **cumulative distribution function** (cdf), that is $P(X \leq x) = F(x)$.

- If X is a random variable with this density function, then for any two real numbers, a and b
$$P(a \leq X \leq b) = \int_{a}^{b} f(x)dx. = F(b) - F(a)$$

Cumulative Distribution Function Properties

Any cdf F has the following properties:

1. F is a non-decreasing function defined on $\mathbb{R}$
2. F is right-continuous, meaning for each a, $F(a) = F(a+) = \lim_{x \to a^+} F(x)$
3. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
4. $P(a < X \leq b) = F(b) - F(a)$ for all real a and b, where $a < b$.
5. $P(X > a) = 1 - F(a)$.
6. $P(X < b) = F(b-) = \lim_{x \to b^-} F(x)$.
7. $P(a < X < b) = F(b-) - F(a)$.
8. $P(X = b) = F(b) - F(b-)$.

For continuous,

$P(X<b) = P(X \leq b) = F(b)$

Quantiles

Let F be a given cumulative distribution and let p be any real number between 0 and 1. The **(100p)th percentile** of the distribution of a continuous random variable X is defined as

$$F^{-1}(p) = min\{x|F(x) \geq p\}.$$

For continuous distributions, $F^{-1}(p)$ is the smallest number x such that $F(x) = p$.

Expected Values for Continuous Random Variables

The **expected** or **mean value** of a continuous random variable X with pdf $f(x)$ is

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

More generally, if h is a function defined on the range of X,

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

The Uniform Distribution

Let $X \sim \text{Unif}(a, b)$

- The pdf of X is:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

- The cdf of X is:

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b < x \end{cases}$$

- $\mu = E(X) = \dfrac{a+b}{2}$
- $\sigma^2 = \text{Var}(X) = \dfrac{(b-a)^2}{12}$

The Exponential Distribution

Let $X \sim \text{Exp}(\lambda)$

- The pdf of X is:

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

- The cdf of X is:

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

- $\mu = E(X) = \dfrac{1}{\lambda}$
- $\sigma^2 = \text{Var}(X) = \dfrac{1}{\lambda^2}$

The Gamma Function

The gamma function $\Gamma(\alpha)$ is defined by:

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

The most important properties of the gamma function are the following:

1. For any $\alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
2. For any positive integer, n, $\Gamma(n) = (n-1)!$
3. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

The Gamma Distribution

Let $X \sim \text{Gamma}(\alpha, \beta)$

- The pdf of X is:

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

- α is the **shape parameter** and β is the **scale parameter**
- $\mu = E(X) = \alpha\beta$
- $\sigma^2 = \text{Var}(X) = \alpha\beta^2$

PDF of a Normal Distribution

A continuous random variable X is said to have a **Normal distribution** with parameters μ and σ (or μ and σ^2), where $-\infty < \mu < \infty$ and $0 < \sigma$, if the pdf of X is:

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

For all $-\infty < x < \infty$.

The cdf of X when $X \sim N(\mu, \sigma)$ is:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\mu)^2/2\sigma^2} dt$$

Standard Normal Distribution

When $X \sim N(\mu, \sigma)$, we can standardize the values by forming:

$$Z = \frac{X - \mu}{\sigma}$$

where $\mu_Z = 0$ and $\sigma_Z = 1$ to get the pdf:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

The cdf of $Z \sim N(0, 1)$ is

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \phi(t)dt = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

Normal Approximation to Binomial

Let X be a binomial random variable based on n trials with success probability p. Then if the binomial probability histogram is not too skewed, X has an approximate Normal distribution with $\mu = np$ and $\sigma = \sqrt{np(1-p)}$. In particular, for x = a possible value of X,

$$\begin{aligned} P(X \leq x) &= Binom(x; n, p) \\ &\approx \text{(area under the normal curve to the left of } x + 0.5) \\ &= \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

In practice, the approximation is adequate provided that both $np \geq 10$ and $n(1-p) \geq 10$.

Using R and R-Studio

1. Download R from https://cran.r-project.org/
2. Download R-Studio from https://www.rstudio.com/