ENE 3031
Computer Simulation
Week 2: Probability and Statistics Review
(attributed by Dr. David Goldsman)

Chuljin Park
Assistant Professor
Industrial Engineering
Hanyang University

Outline

Preliminaries

Will assume that you know about sample spaces, events, and the definition of probability.

Definition: $P(A|B) \equiv P(A \cap B)/P(B)$ is the *conditional probability of A given B*.

Example: Toss a fair die. Let $A = \{1, 2, 3\}$ and $B = \{3, 4, 5, 6\}$. Then

$$P(A|B) \;=\; \frac{P(A \cap B)}{P(B)} \;=\; \frac{1/6}{4/6} \;=\; 1/4. \quad \Box$$

Definition: If $P(A \cap B) = P(A)P(B)$, then A and B are *independent* events.

Theorem: If A and B are independent, then $P(A|B) = P(A)$.

Example: Toss two dice. Let $A =$ “Sum is 7” and $B =$ “First die is 4”. Then

$$P(A) = 1/6, \quad P(B) = 1/6, \quad \text{and}$$

$$P(A \cap B) = P((4, 3)) = 1/36 = P(A)P(B).$$

So A and B are independent. $\Box$

Definition: A *random variable* (RV) X is a function from the sample space Ω to the real line, i.e., $X : \Omega \to \mathsf{R}$.

Example: Let X be the sum of two dice rolls. Then $X((4,6)) = 10$. In addition,

$$P(X = x) \;=\; \begin{cases} 1/36 & \text{if } x = 2 \\ 2/36 & \text{if } x = 3 \\ \vdots & \\ 1/36 & \text{if } x = 12 \\ 0 & \text{otherwise} \end{cases} \qquad \square$$

Definition: If the number of possible values of a RV X is finite or countably infinite, then X is a *discrete* RV. Its *probability mass function* (pmf) is $f(x) \equiv P(X = x)$. Note that $\sum_x f(x) = 1$.

Example: Flip 2 coins. Let X be the number of heads.

$$f(x) \;=\; \begin{cases} 1/4 & \text{if } x = 0 \text{ or } 2 \\ 1/2 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases} \qquad \square$$

Examples: Here are some well-known discrete RV's that you may know: Bernoulli(p), Binomial(n,p), Geometric(p), Negative Binomial, Poisson(λ), etc.

Definition: A *continuous* RV is one with probability zero at every individual point. A RV is continuous if there exists a *probability density function* (pdf) $f(x)$ such that $P(X \in A) = \int_A f(x)\,dx$ for every set A. Note that $\int_x f(x)\,dx = 1$.

Example: Pick a random number between 3 and 7. Then

$$f(x) \;=\; \begin{cases} 1/4 & \text{if } 3 \le x \le 7 \\ 0 & \text{otherwise} \end{cases} \qquad \square$$

Examples: Here are some well-known continuous RV's: Uniform(a,b), Exponential(λ), Normal(μ, σ^2), etc.

Notation: "$\sim$" means "is distributed as". For instance, $X \sim \mathrm{Unif}(0,1)$ means that X has the uniform distribution on [0,1].

Definition: For any RV X (discrete or continuous), the *cumulative distribution function* (cdf) is

$$F(x) \;\equiv\; P(X \le x) \;=\; \begin{cases} \sum_{y \le x} f(y) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{x} f(y)\,dy & \text{if } X \text{ is continuous} \end{cases}$$

Note that $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. In addition, if X is continuous, then $\frac{d}{dx}F(x) = f(x)$.

Example: Flip 2 coins. Let X be the number of heads.

$$F(x) \;=\; \begin{cases} 0 & \text{if } x < 0 \\ 1/4 & \text{if } 0 \le x < 1 \\ 3/4 & \text{if } 1 \le x < 2 \\ 1 & \text{if } x \ge 2 \end{cases} \qquad \square$$

Example: if $X \sim \mathrm{Exp}(\lambda)$ (i.e., X is exponential with parameter λ), then $f(x) = \lambda e^{-\lambda x}$ and $F(x) = 1 - e^{-\lambda x}$, $x \ge 0$. $\square$

Outline

1. Preliminaries
2. Simulating Random Variables
3. Great Expectations
4. Functions of a Random Variable
5. Jointly Distributed Random Variables
6. Covariance and Correlation
7. Some Probability Distributions
8. Limit Theorems
9. Statistics Tidbits

Simulating Random Variables

We'll make a brief aside here to show how to simulate some very simple random variables.

Example (Discrete Uniform): Consider a D.U. on $\{1, 2, \ldots, n\}$, i.e., $X = i$ with probability $1/n$ for $i = 1, 2, \ldots, n$. (Think of this as an n-sided dice toss for you Dungeons and Dragons fans.)

If $U \sim \text{Unif}(0, 1)$, we can obtain a D.U. random variate simply by setting $X = \lceil nU \rceil$, where $\lceil \cdot \rceil$ is the "ceiling" (or "round up") function.

For example, if $n = 10$ and we sample a Unif(0,1) random variable $U = 0.73$, then $X = \lceil 7.3 \rceil = 8$. $\Box$

Example (Another Discrete Random Variable):

$$P(X = x) = \begin{cases} 0.25 & \text{if } x = -2 \\ 0.10 & \text{if } x = 3 \\ 0.65 & \text{if } x = 4.2 \\ 0 & \text{otherwise} \end{cases}$$

Can't use a die toss to simulate this random variable. Instead, use what's called the *inverse transform method*.

x	$f(x)$	$P(X \le x)$	Unif(0,1)'s
-2	0.25	0.25	$[0.00, 0.25]$
3	0.10	0.35	$(0.25, 0.35]$
4.2	0.65	1.00	$(0.35, 1.00)$

Sample $U \sim \text{Unif}(0, 1)$. Choose the corresponding x-value, i.e., $X = F^{-1}(U)$. For example, $U = 0.46$ means that $X = 4.2$. $\Box$

Now we'll use the *inverse transform method* to generate a continuous random variable. We'll talk about the following result a little later...

Theorem: If X is a continuous random variable with cdf $F(x)$, then the random variable $F(X) \sim \mathsf{Unif}(0, 1)$.

This suggests a way to generate realizations of the RV X. Simply set $F(X) = U \sim \mathsf{Unif}(0, 1)$ and solve for $X = F^{-1}(U)$.

Example: Suppose $X \sim \text{Exp}(\lambda)$. Then $F(x) = 1 - e^{-\lambda x}$ for $x > 0$. Set $F(X) = 1 - e^{-\lambda X} = U$. Solve for X,

$$X = \frac{-1}{\lambda} \ell\text{n}(1 - U) \sim \text{Exp}(\lambda). \quad \Box$$

Example (Generating Uniforms): All of the above RV generation examples relied on our ability to generate a Unif(0,1) RV. For now, let's assume that we can generate numbers that are "practically" iid Unif(0,1).

If you don't like programming, you can use Excel function `RAND()` or something similar to generate Unif(0,1)'s.

Here's an algorithm to generate *pseudo-random numbers (PRN's)*, i.e., a series $R_1, R_2, \ldots$ of *deterministic* numbers that *appear* to be iid Unif(0,1). Pick a *seed* integer X_0, and calculate

$$X_i \ = \ 16807X_{i-1}\text{mod}(2^{31}-1), \quad i = 1, 2, \ldots.$$

Then set $R_i = X_i/(2^{31}-1)$, $i = 1, 2, \ldots$.

Here's an easy FORTRAN implementation of the above algorithm (from Bratley, Fox, and Schrage).

```
FUNCTION UNIF(IX)
K1 = IX/127773     (this division truncates, e.g., 5/3 = 1.)
IX = 16807*(IX - K1*127773) - K1*2836     (update seed)
IF(IX.LT.0)IX = IX + 2147483647
UNIF = IX * 4.656612875E-10
RETURN
END
```

In the above function, we input a positive integer IX and the function returns the PRN UNIF, as well as an updated IX that we can use again. □

Some Exercises: In the following, I'll assume that you can use Excel (or whatever) to simulate independent Unif(0,1) RV's. (We'll review independence in a little while.)

1. Make a histogram of $X_i = -\ell\text{n}(U_i)$, for $i = 1, 2, \ldots, 10000$, where the U_i's are independent Unif(0,1) RV's. What kind of distribution does it look like?
2. Suppose X_i and Y_i are independent Unif(0,1) RV's, $i = 1, 2, \ldots, 10000$. Let $Z_i = \sqrt{-2\ell\text{n}(X_i)}\ \sin(2\pi Y_i)$, and make a histogram of the Z_i's based on the 10000 replications.
3. Suppose X_i and Y_i are independent Unif(0,1) RV's, $i = 1, 2, \ldots, 10000$. Let $Z_i = X_i/(X_i - Y_i)$, and make a histogram of the Z_i's based on the 10000 replications. This may be somewhat interesting. It's possible to derive the distribution analytically, but it takes a lot of work.

Outline

Great Expectations

Definition: The *expected value* (or *mean*) of a RV X is

$$\mathrm{E}[X] \equiv \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x f(x)\,dx & \text{if } X \text{ is continuous} \end{cases} = \int_{\mathbb{R}} x\,dF(x).$$

Example: Suppose that $X \sim \mathrm{Bernoulli}(p)$. Then

$$X = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p\ (=q) \end{cases}$$

and we have $\mathrm{E}[X] = \sum_x x f(x) = p$. $\square$

Example: Suppose that $X \sim \mathrm{Uniform}(a,b)$. Then

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

and we have $\mathrm{E}[X] = \int_{\mathbb{R}} x f(x)\,dx = (a+b)/2$. $\square$

"Definition" (the "Law of the Unconscious Statistician"): Suppose that $h(X)$ is some function of the RV X. Then

$$\mathrm{E}[h(X)] = \begin{cases} \sum_x h(x) f(x) & \text{if } X \text{ is disc} \\ \int_{\mathbb{R}} h(x) f(x)\,dx & \text{if } X \text{ is cts} \end{cases} = \int_{\mathbb{R}} h(x)\,dF(x).$$

Example: Suppose X is the following discrete RV:

x	2	3	4
$f(x)$	0.3	0.6	0.1

Then $\mathrm{E}[X^3] = \sum_x x^3 f(x) = 8(0.3) + 27(0.6) + 64(0.1) = 25$. $\square$

Example: Suppose $X \sim \mathrm{U}(0,2)$. Then

$$\mathrm{E}[X^n] = \int_{\mathbb{R}} x^n f(x)\,dx = 2^n/(n+1). \quad \square$$

Definitions: $\mathrm{E}[X^n]$ is the nth *moment* of X. $\mathrm{E}[(X-\mathrm{E}[X])^n]$ is the nth *central moment* of X.

$\mathrm{Var}(X) \equiv \mathrm{E}[(X-\mathrm{E}[X])^2] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$ is the *variance* of X. The *standard deviation* of X is $\sqrt{\mathrm{Var}(X)}$.

Example: Suppose $X \sim \mathrm{Bern}(p)$. Recall that $\mathrm{E}[X] = p$. Then

$$\mathrm{E}[X^2] = \sum_x x^2 f(x) = p \quad \text{and}$$

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = p(1-p). \quad \square$$

Example: Suppose $X \sim \mathrm{U}(0,2)$. By previous examples, $\mathrm{E}[X] = 1$ and $\mathrm{E}[X^2] = 4/3$. So

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = 1/3. \quad \square$$

Theorem: $\mathrm{E}[aX+b] = a\mathrm{E}[X] + b$ and $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$.

Definition: $M_X(t) \equiv \mathrm{E}[e^{tX}]$ is the *moment generating function* (mgf) of the RV X. ($M_X(t)$ is a function of t, *not* of X!)

Example: $X \sim \mathrm{Bern}(p)$. Then

$$M_X(t) = \mathrm{E}[e^{tX}] = \sum_x e^{tx} f(x) = e^{t\cdot 1}p + e^{t\cdot 0}q = pe^t + q. \quad \square$$

Example: $X \sim \mathrm{Exp}(\lambda)$. Then

$$M_X(t) = \int_{\Re} e^{tx} f(x)\,dx = \lambda \int_0^\infty e^{(t-\lambda)x}\,dx = \frac{\lambda}{\lambda - t} \quad \text{if } \lambda > t. \quad \square$$

Theorem: Under certain technical conditions,

$$\mathrm{E}[X^k] = \frac{d^k}{dt^k} M_X(t)\Big|_{t=0}, \quad k = 1, 2, \ldots.$$

Thus, you can *generate* the moments of X from the mgf.

Example: $X \sim \text{Exp}(\lambda)$. Then $M_X(t) = \frac{\lambda}{\lambda - t}$ for $\lambda > t$. So

$$\text{E}[X] \;=\; \frac{d}{dt}M_X(t)\Big|_{t=0} \;=\; \frac{\lambda}{(\lambda-t)^2}\Big|_{t=0} \;=\; 1/\lambda.$$

Further,

$$\text{E}[X^2] \;=\; \frac{d^2}{dt^2}M_X(t)\Big|_{t=0} \;=\; \frac{2\lambda}{(\lambda-t)^3}\Big|_{t=0} \;=\; 2/\lambda^2.$$

Thus,

$$\text{Var}(X) \;=\; \text{E}[X^2] - (\text{E}[X])^2 \;=\; \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \;=\; 1/\lambda^2. \quad \square$$

Moment generating functions have many other important uses, some of which we'll talk about in this course.

Outline

Functions of a Random Variable

Problem: Suppose we have a RV X with pdf/pmf $f(x)$. Let $Y = h(X)$. Find $g(y)$, the pdf/pmf of Y.

Discrete Example: Let X denote the number of H's from two coin tosses. We want the pmf for $Y = X^3 - X$.

x	0	1	2
$f(x)$	1/4	1/2	1/4
$y = x^3 - x$	0	0	6

This implies that $g(0) = P(Y=0) = P(X = 0 \text{ or } 1) = 3/4$ and $g(6) = P(Y=6) = 1/4$. In other words,

$$g(y) \;=\; \begin{cases} 3/4 & \text{if } y = 0 \\ 1/4 & \text{if } y = 6 \end{cases}. \quad \square$$

Continuous Example: Suppose X has pdf $f(x) = |x|$, $-1 \le x \le 1$. Find the pdf of $Y = X^2$.

First of all, the cdf of Y is

$$\begin{aligned} G(y) &= P(Y \le y) \\ &= P(X^2 \le y) \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} |x|\,dx \;=\; y, \quad 0 < y < 1. \end{aligned}$$

Thus, the pdf of Y is $g(y) = G'(y) = 1$, $0 < y < 1$, indicating that $Y \sim \text{Unif}(0,1)$. $\square$

Inverse Transform Theorem: Suppose X is a continuous random variable having cdf $F(x)$. Then, amazingly, $F(X) \sim \text{Unif}(0,1)$.

Proof: Let $Y = F(X)$. Then the cdf of Y is

$$\begin{aligned} P(Y \le y) &= P(F(X) \le y) \\ &= P(X \le F^{-1}(y)) \\ &= F(F^{-1}(y)) \; = \; y, \end{aligned}$$

which is the cdf of the Unif(0,1). $\square$

This result is of fundamental importance when it comes to generating random variates during a simulation.

Example: Suppose $X \sim \text{Exp}(\lambda)$, so that its cdf is $F(x) = 1 - e^{-\lambda x}$ for $x \ge 0$.

Then the Inverse Transform Theorem implies that

$$F(X) \; = \; 1 - e^{-\lambda X} \; \sim \; \text{Unif}(0,1).$$

Let $U \sim \text{Unif}(0,1)$ and set $F(X) = U$.

After a little algebra, we find that

$$X \; = \; \frac{-1}{\lambda} \, \ell\text{n}(1-U) \; \sim \; \text{Exp}(\lambda).$$

This is how you can generate exponential random variates. $\square$

Exercise: Suppose that X has the Weibull distribution with cdf

$$F(x) = 1 - e^{-(\lambda x)^\beta}, x > 0.$$

If you set $F(X) = U$ and solve for X, show that you get

$$X = \frac{1}{\lambda} [-\ell\text{n}(1-U)]^{1/\beta}.$$

Now pick your favorite λ and β, and use this result to generate values of X. In fact, make a histogram of your X values. Are there any interesting values of λ and β you could've chosen?

Outline

Jointly Distributed Random Variables

Consider two random variables interacting together — think height and weight.

Definition: The *joint cdf* of X and Y is

$$F(x,y) \;\equiv\; P(X \le x, Y \le y), \quad \text{for all } x, y.$$

Remark: The *marginal cdf* of X is $F_X(x) = F(x, \infty)$. (We use the X subscript to remind us that it's just the cdf of X all by itself.) Similarly, the *marginal cdf* of Y is $F_Y(y) = F(\infty, y)$.

Definition: If X and Y are discrete, then the *joint pmf* of X and Y is $f(x,y) \equiv P(X = x, Y = y)$. Note that $\sum_x \sum_y f(x,y) = 1$.

Remark: The *marginal pmf* of X is

$$f_X(x) \;=\; P(X = x) \;=\; \sum_y f(x,y).$$

The *marginal pmf* of Y is

$$f_Y(y) \;=\; P(Y = y) \;=\; \sum_x f(x,y).$$

Example: The following table gives the joint pmf $f(x,y)$, along with the accompanying marginals.

	$X=2$	$X=3$	$X=4$	$f_Y(y)$
$Y=4$	0.3	0.2	0.1	0.6
$Y=6$	0.1	0.2	0.1	0.4
$f_X(x)$	0.4	0.4	0.2	1

Definition: If X and Y are continuous, then the *joint pdf* of X and Y is $f(x,y) \equiv \frac{\partial^2}{\partial x \partial y} F(x,y)$. Note that $\int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y)\,dx\,dy = 1$.

Remark: The *marginal pdf's* of X and Y are

$$f_X(x) \;=\; \int_{\mathbb{R}} f(x,y)\,dy \quad \text{and} \quad f_Y(y) \;=\; \int_{\mathbb{R}} f(x,y)\,dx.$$

Example: Suppose the joint pdf is

$$f(x,y) \;=\; \frac{21}{4} x^2 y, \quad x^2 \le y \le 1.$$

Then the marginal pdf's are:

Definition: X and Y are *independent* RV's if

$$f(x,y) \;=\; f_X(x) f_Y(y) \quad \text{for all } x, y.$$

Theorem: X and Y are indep if you can write their joint pdf as $f(x,y) = a(x)b(y)$ for some functions $a(x)$ and $b(y)$, and x and y don't have funny limits (their domains do not depend on each other).

Examples: If $f(x,y) = cxy$ for $0 \le x \le 2$, $0 \le y \le 3$, then X and Y are .

If $f(x,y) = \frac{21}{4} x^2 y$ for $x^2 \le y \le 1$, then X and Y are .

If $f(x,y) = c/(x+y)$ for $1 \le x \le 2$, $1 \le y \le 3$, then X and Y are . □

Definition: The *conditional pdf* (or *pmf*) of Y given $X = x$ is $f(y|x) \equiv f(x,y)/f_X(x)$.

Example: Suppose $f(x,y) = \frac{21}{4}x^2 y$ for $x^2 \le y \le 1$. Then

$$f(y|x) = \boxed{} \quad \square$$

Theorem: If X and Y are indep, then $f(y|x) = f_Y(y)$ for all x, y.

Definition: The *conditional expectation* of Y given $X = x$ is

$$\mathrm{E}[Y|X=x] \equiv \begin{cases} \sum_y y f(y|x) & \text{discrete} \\ \int_{\mathbb{R}} y f(y|x)\, dy & \text{continuous} \end{cases}$$

Old Cts Example: $f(x,y) = \frac{21}{4}x^2 y$, if $x^2 \le y \le 1$. Then

$$\mathrm{E}[Y|x] = \boxed{} \quad \square$$

Theorem (double expectations): $\mathrm{E}[\mathrm{E}(Y|X)] = \mathrm{E}[Y]$.

Proof (cts case): By the Unconscious Statistician,

$$\begin{aligned} \mathrm{E}[\mathrm{E}(Y|X)] &= \int_{\mathbb{R}} \mathrm{E}(Y|x) f_X(x)\, dx \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} y f(y|x)\, dy \right) f_X(x)\, dx \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} y f(y|x) f_X(x)\, dx\, dy \\ &= \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x,y)\, dx\, dy \\ &= \int_{\mathbb{R}} y f_Y(y)\, dy = \mathrm{E}[Y]. \quad \square \end{aligned}$$

Old Example: Suppose $f(x,y) = \frac{21}{4}x^2 y$, if $x^2 \le y \le 1$. Note that, by previous examples, we know $f_X(x)$, $f_Y(y)$, and $\mathrm{E}[Y|x]$.

Solution #1 (old, boring way):

$$\mathrm{E}[Y] = \boxed{}.$$

Solution #2 (new, exciting way):

$$\mathrm{E}[Y] \quad \boxed{}.$$

Notice that both answers are the same (good)! $\square$

Outline

"Definition": Suppose that $h(X,Y)$ is some function of the RV's X and Y. Then

$$\mathrm{E}[h(X,Y)] \;=\; \begin{cases} \sum_x \sum_y h(x,y) f(x,y) & \text{if } (X,Y) \text{ is discrete} \\ \int_{\mathbb{R}} \int_{\mathbb{R}} h(x,y) f(x,y)\, dx\, dy & \text{if } (X,Y) \text{ is continuous} \end{cases}$$

Theorem: Whether or not X and Y are independent, we have $\mathrm{E}[X+Y] = \boxed{}$.

Theorem: If X and Y are *independent*, then $\mathrm{Var}(X+Y) = \boxed{}$.

(Stay tuned for dependent case.)

Definition: $X_1, \ldots, X_n$ form a *random sample* from $f(x)$ if (i) $X_1, \ldots, X_n$ are independent, and (ii) each X_i has the same pdf (or pmf) $f(x)$.

Notation: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x)$. (The term "iid" reads *independent and identically distributed*.)

Example: If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x)$ and the *sample mean* $\bar{X}_n \equiv \sum_{i=1}^n X_i/n$, then $\mathrm{E}[\bar{X}_n] = \mathrm{E}[X_i]$ and $\mathrm{Var}(\bar{X}_n) = \mathrm{Var}(X_i)/n$. Thus, the variance *decreases* as n increases. $\square$

But not all RV's are independent...

Covariance and Correlation

Definition: The *covariance* between X and Y is

$$\mathrm{Cov}(X,Y) \;\equiv\; \mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])] \;=\; \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y].$$

Note that $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$.

Theorem: If X and Y are independent RV's, then $\mathrm{Cov}(X,Y) = 0$.

Remark: $\mathrm{Cov}(X,Y) = 0$ doesn't mean X and Y are independent!

Example: Suppose $X \sim \mathrm{Unif}(-1,1)$ and $Y = X^2$. Then X and Y are clearly dependent. However,

$$\mathrm{Cov}(X,Y) \;=\; \boxed{} \qquad \square$$

Theorem: $\mathrm{Cov}(aX, bY) = ab\mathrm{Cov}(X,Y)$.

Theorem: Whether or not X and Y are independent,

$$\mathrm{Var}(X+Y) \;=\; \boxed{}$$

and

$$\mathrm{Var}(X-Y) \;=\; \boxed{}.$$

Definition: The *correlation* between X and Y is

$$\rho \;\equiv\; \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}.$$

Theorem: $-1 \le \rho \le 1$.

Example: Consider the following joint pmf.

$f(x,y)$	$X=2$	$X=3$	$X=4$	$f_Y(y)$
$Y=40$	0.00	0.20	0.10	0.3
$Y=50$	0.15	0.10	0.05	0.3
$Y=60$	0.30	0.00	0.10	0.4
$f_X(x)$	0.45	0.30	0.25	1

$\mathrm{E}[X]=2.8$, $\mathrm{Var}(X)=0.66$, $\mathrm{E}[Y]=51$, $\mathrm{Var}(Y)=69$,

$$\mathrm{E}[XY] = \boxed{},$$

and

$$\rho = \boxed{}. \quad \square$$

Outline

Some Probability Distributions

First, some discrete distributions…

$X \sim \mathrm{Bernoulli}(p)$.

$$f(x) = \begin{cases} p & \text{if } x=1 \\ 1-p\ (=q) & \text{if } x=0 \end{cases}$$

$\mathrm{E}[X]=p$, $\mathrm{Var}(X)=pq$.

$Y \sim \mathrm{Binomial}(n,p)$. If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{Bern}(p)$ (i.e., *Bernoulli(p) trials*), then $Y = \sum_{i=1}^n X_i \sim \mathrm{Bin}(n,p)$.

$$f(y) = \binom{n}{y} p^y q^{n-y}, \quad y = 0,1,\ldots,n.$$

$\mathrm{E}[Y]=np$, $\mathrm{Var}(Y)=npq$.

$X \sim \mathrm{Geometric}(p)$ is the number of Bern(p) trials until a success occurs. For example, "FFFS" implies that $X=4$.

$$f(x) = q^{x-1}p, \quad x = 1,2,\ldots.$$

$\mathrm{E}[X]=1/p$, $\mathrm{Var}(X)=q/p^2$.

$Y \sim \mathrm{NegBin}(r,p)$ is the sum of r iid Geom(p) RV's, i.e., the time until the rth success occurs. For example, "FFFSSFS" implies that $\mathrm{NegBin}(3,p)=7$.

$$f(y) = \binom{y-1}{r-1} q^{y-r}p^r, \quad y = r, r+1, \ldots.$$

$\mathrm{E}[Y]=r/p$, $\mathrm{Var}(Y)=qr/p^2$.

$X \sim \text{Poisson}(\lambda)$.

Definition: A *counting process* $N(t)$ tallies the number of "arrivals" observed in $[0, t]$. A *Poisson process* is a counting process satisfying the following.

i. Arrivals occur one-at-a-time at rate λ (e.g., $\lambda = 4$ customers/hr)
ii. Independent increments, i.e., the numbers of arrivals in disjoint time intervals are independent.
iii. Stationary increments, i.e., the distribution of the number of arrivals in $[s, s+t]$ only depends on t.

$X \sim \text{Pois}(\lambda)$ is the number of arrivals that a Poisson process experiences in one time unit, i.e., $N(1)$.

$$f(x) \;=\; \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots.$$

$\text{E}[X] = \lambda = \text{Var}(X)$.

Now, some continuous distributions...

$X \sim \text{Uniform}(a, b)$. $f(x) = \frac{1}{b-a}$ for $a \le x \le b$, $\text{E}[X] = \frac{a+b}{2}$, and $\text{Var}(X) = \frac{(b-a)^2}{12}$.

$X \sim \text{Exponential}(\lambda)$. $f(x) = \lambda e^{-\lambda x}$ for $x \ge 0$, $\text{E}[X] = 1/\lambda$, and $\text{Var}(X) = 1/\lambda^2$.

Theorem: The exponential distribution has the *memoryless property* (and is the only continuous distribution with this property), i.e., for $s, t > 0$, $P(X > s+t | X > s) = \boxed{}$.

Example: Suppose $X \sim \text{Exp}(\lambda = 1/100)$. Then

$$P(X > 200 | X > 50) \;=\; \boxed{}. \quad \Box$$

$X \sim \text{Gamma}(\alpha, \lambda)$.

$$f(x) \;=\; \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x \ge 0,$$

where the gamma function is

$$\Gamma(\alpha) \;\equiv\; \int_0^\infty t^{\alpha-1} e^{-t}\, dt.$$

$\text{E}[X] = \alpha/\lambda$, $\text{Var}(X) = \alpha/\lambda^2$.

If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$, then $Y \equiv \sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda)$. The $\text{Gamma}(n, \lambda)$ is also called the $\text{Erlang}_n(\lambda)$. It has cdf

$$F_Y(y) \;=\; 1 - e^{-\lambda y} \sum_{j=0}^{n-1} \frac{(\lambda y)^j}{j!}, \quad y \ge 0.$$

$X \sim \text{Triangular}(a, b, c)$. Good for modeling things with limited data — a is the smallest possible value, b is the "most likely," and c is the largest.

$$f(x) \;=\; \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{if } a < x \le b \\ \frac{2(c-x)}{(c-b)(c-1)} & \text{if } b < x \le c \\ 0 & \text{otherwise} \end{cases}.$$

$\text{E}[X] = (a+b+c)/3$.

$X \sim \text{Beta}(a, b)$. $f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$ for $0 \le x \le 1$ and $a, b > 0$.

$$\text{E}[X] \;=\; \frac{a}{a+b} \quad \text{and} \quad \text{Var}(X) \;=\; \frac{ab}{(a+b)^2(a+b+1)}.$$

$X \sim \text{Normal}(\mu, \sigma^2)$. Most important distribution.

$$f(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \quad x \in \mathsf{R}.$$

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

Theorem: If $X \sim \text{Nor}(\mu, \sigma^2)$, then $aX + b \sim \text{Nor}(a\mu + b, a^2\sigma^2)$.

Corollary: If $X \sim \text{Nor}(\mu, \sigma^2)$, then $Z \equiv \frac{X-\mu}{\sigma} \sim \text{Nor}(0,1)$, the *standard normal distribution*, with pdf $\phi(z) \equiv \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$ and cdf $\Phi(z)$, which is tabled. E.g., $\Phi(1.96) \doteq 0.975$.

Theorem: If X_1 and X_2 are *independent* with $X_i \sim \text{Nor}(\mu_i, \sigma_i^2)$, $i = 1, 2$, then $X_1 + X_2 \sim \text{Nor}(\boxed{}, \boxed{})$.

Example: Suppose $X \sim \text{Nor}(3,4)$, $Y \sim \text{Nor}(4,6)$, and X and Y are independent. Then $2X - 3Y + 1 \sim \boxed{}$. $\square$

There are a number of distributions (including the normal) that come up in statistical sampling problems. Here are a few:

Definitions: If $Z_1, Z_2, \ldots, Z_k$ are iid Nor(0,1), then $Y = \sum_{i=1}^k Z_i^2$ has the χ^2 *distribution with k degrees of freedom (df)*. Notation: $Y \sim \chi^2(k)$. Note that $\mathrm{E}[Y] = k$ and $\mathrm{Var}(Y) = 2k$.

If $Z \sim \text{Nor}(0,1)$, $Y \sim \chi^2(k)$, and Z and Y are independent, then $T = Z/\sqrt{Y/k}$ has the *Student t distribution with k df*. Notation: $T \sim t(k)$. Note that the $t(1)$ is the *Cauchy* distribution.

If $Y_1 \sim \chi^2(m)$, $Y_2 \sim \chi^2(n)$, and Y_1 and Y_2 are independent, then $F = (Y_1/m)/(Y_2/n)$ has the *F distribution with m and n df.* Notation: $F \sim F(m,n)$.

Outline

Limit Theorems

Corollary (of theorem from previous section): If $X_1, \ldots, X_n$ are iid $\text{Nor}(\mu, \sigma^2)$, then the sample mean $\bar{X}_n \sim \text{Nor}(\mu, \sigma^2/n)$.

This is a special case of the *Law of Large Numbers*, which says that $\bar{X}_n$ approximates μ well as n becomes large.

Definition: The sequence of RV's $Y_1, Y_2, \ldots$ with respective cdf's $F_{Y_1}(y), F_{Y_2}(y), \ldots$ *converges in distribution* to the RV Y having cdf $F_Y(y)$ if $\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)$ for all y belonging to the continuity set of Y. Notation: $Y_n \xrightarrow{d} Y$.

Idea: If $Y_n \xrightarrow{d} Y$ and n is large, then you ought to be able to approximate the distribution of Y_n by the limit distribution of Y.

Central Limit Theorem: If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} f(x)$ with mean μ and variance σ^2, then

$$Z_n \equiv \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \text{Nor}(0,1).$$

Thus, the cdf of Z_n approaches $\Phi(z)$ as n increases. The CLT usually works well if the pdf/pmf is fairly symmetric and $n \geq 15$.

Example: Suppose $X_1, X_2, \ldots, X_{100} \overset{\text{iid}}{\sim} \mathsf{Exp}(1)$ (so $\mu = \sigma^2 = 1$).

$$P\left(90 \leq \sum_{i=1}^{100} X_i \leq 110\right) = \boxed{} \approx$$

Exercise: Demonstrate that the CLT actually works.

1. Pick your favorite RV X_1. Simulate it and make a histogram.
2. Now suppose X_1 and X_2 are iid from your favorite distribution. Make a histogram of $X_1 + X_2$.
3. Now $X_1 + X_2 + X_3$.
4. ... Now $X_1 + X_2 + \cdots + X_n$ for some reasonably large n.
5. Does the CLT work for the Cauchy distribution, i.e., $X = \tan(2\pi U)$, where $U \sim \text{Unif}(0,1)$?

Outline

Statistics Tidbits

For now, suppose that $X_1, X_2, \ldots, X_n$ are iid from some distribution with finite mean μ and finite variance σ^2.

For this iid case, we have already seen that $\text{E}[\bar{X}_n] = \mu$, i.e., $\bar{X}_n$ is *unbiased* for μ.

Definition: The *sample variance* is $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$.

Theorem: If $X_1, X_2, \ldots, X_n$ are iid with variance σ^2, then $\text{E}[S^2] = \sigma^2$, i.e., S^2 is unbiased for σ^2.

If $X_1, X_2, \ldots, X_n$ are iid $\text{Nor}(\mu, \sigma^2)$, then $\bar{X}_n \sim \text{Nor}(\mu, \sigma^2/n)$ and $S^2 \sim \frac{\sigma^2\chi^2(n-1)}{n-1}$ (and $\bar{X}_n$ and S^2 are independent).

These facts can be used to construct *confidence intervals* (CIs) for μ and σ^2 under a variety of assumptions.

A $100(1-\alpha)\%$ two-sided CI for an unknown parameter θ is a random interval $[L, U]$ such that $P(L \le \theta \le U) = 1 - \alpha$.

Here are some examples/theorems, all of which assume that the X_i's are iid normal…

Example: If σ^2 is *known*, then a $100(1-\alpha)\%$ CI for μ is

$$\bar{X}_n - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \;\le\; \mu \;\le\; \bar{X}_n + z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}},$$

where z_γ is the $1-\gamma$ quantile of the standard normal distribution, i.e., $z_\gamma \equiv \Phi^{-1}(1-\gamma)$.

Example: If σ^2 is *unknown*, then a $100(1-\alpha)\%$ CI for μ is

$$\bar{X}_n - t_{\alpha/2,n-1}\sqrt{\frac{S^2}{n}} \;\le\; \mu \;\le\; \bar{X}_n + t_{\alpha/2,n-1}\sqrt{\frac{S^2}{n}},$$

where $t_{\gamma,\nu}$ is the $1-\gamma$ quantile of the $t(\nu)$ distribution.

Example: A $100(1-\alpha)\%$ CI for σ^2 is

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \;\le\; \sigma^2 \;\le\; \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}},$$

where $\chi^2_{\gamma,\nu}$ is the $1-\gamma$ quantile of the $\chi^2(\nu)$ distribution.