Lecture 16: Midterm 2 Review

Sta230 / Mth 230

Colin Rundel

March 26, 2014

Properties of Expected Value

- $E(c) = c$
- $E(I_A) = P(A)$
- $E[g(X)] = \sum_{\text{all } x} g(x)\ P(X = x)$
- $E(cX) = cE(x)$
- $E(X + Y) = E(X) + E(Y)$
- $E(XY) = E(X)E(Y)$ if X and Y are independent.

Properties of Variance

- $Var(aX) = a^2 Var(X)$
- $Var(X + c) = Var(X)$
- $Var(c) = 0$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$
- $Var(aX + bY + c) = a^2\ Var(X) + b^2\ Var(Y) + 2ab\ Cov(X, Y)$
- $Var\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n \sum_{j=1}^n Cov(c_i X_i, c_j X_j)$
  $= \sum_{i=1}^n c_i^2 Var(X_i) + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n c_i c_j Cov(X_i, X_j)$

Properties of Covariance

- $Cov(X, Y) = Cov(Y, X)$
- $Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)] = E(XY) - \mu_x \mu_y$
- $Cov(X, Y) = 0$ if X and Y are independent
- $Cov(X, c) = 0$
- $Cov(X, X) = Var(X)$
- $Cov(aX, bY) = ab\ Cov(X, Y)$
- $Cov(X + a, Y + b) = Cov(X, Y)$

Moments

Raw moment:

$$\mu'_n = E(X^n)$$

Central moment:

$$\mu_n = E[(X-\mu)^2]$$

Normalized / Standardized moment:

$$\frac{\mu_n}{\sigma^n}$$

Moment Generating Function

The moment generating function of a random variable X is defined for all real values of t by

$$M_X(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} P(X=x) & \text{If } X \text{ is discrete} \\ \int_x e^{tx} P(X=x)dx & \text{If } X \text{ is continuous} \end{cases}$$

This is called the moment generating function because we can obtain the raw moments of X by successively differentiating $M_X(t)$ and evaluating at $t=0$.

$$M_X(0) = E[e^0] = 1 = \mu'_0$$

$$M'_X(t) = \frac{d}{dt}E[e^{tX}] = E\left[\frac{d}{dt}e^{tX}\right] = E[Xe^{tX}]$$

$$M'_X(0) = E[Xe^0] = E[X] = \mu'_1$$

$$M''_X(t) = \frac{d}{dt}M'_X(t) = \frac{d}{dt}E[Xe^{tX}] = E\left[\frac{d}{dt}(Xe^{tX})\right] = E[X^2e^{tX}]$$

$$M''_X(0) = E[X^2e^0] = E[X^2] = \mu'_2$$

Moment Generating Function - Properties

If X and Y are independent random variables then the moment generating function for the distribution of $X+Y$ is

$$M_{X+Y}(t) = E[e^{t(X+Y)}] = E[e^{tX}e^{tY}] = E[e^{tX}]E[e^{tY}] = M_X(t)M_Y(t)$$

Similarly, the moment generating function for S_n, the sum of iid random variables $X_1, X_2, \ldots, X_n$ is

$$M_{S_n}(t) = [M_{X_i}(t)]^n$$

Cumulative Distribution Function

We have seen a variety of problems where we find $P(X <= x)$ or $P(X > x)$ etc. The former is given a special name - the cumulative distribution function.

If X is discrete with probability mass function $f(x)$ then

$$P(X \le x) = F(x) = \sum_{z=-\infty}^{x} f(z)$$

If X is continuous with probability density function $f(x)$ then

$$P(X \le x) = F(x) = \int_{-\infty}^{x} f(z)dz$$

CDF is defined for for all $-\infty < x < \infty$ and follows the following rules:

- $\lim_{x \to -\infty} F(x) = 0$
- $\lim_{x \to \infty} F(x) = 1$
- $x < y \Rightarrow F(x) < F(y)$

Probability Density Function

For a continuous probability distribution

$$P(X = x) = 0 \text{ for all } x$$

As such we define the probability density function to be

$$f_X(x) = \lim_{\epsilon \to 0} P(X \in [x, x+\epsilon])/\epsilon$$

A pdf is defined for for all $-\infty < x < \infty$ and follows the following rules:

- $\int_{-\infty}^{\infty} f_X(x)dx = 1$
- $\int_{-\infty}^{x} f_X(t)dt = F_X(x) \;\Leftrightarrow\; f_X(x) = \frac{d}{dx}F_X(x)$
- $f_X(x) \geq 0$ for all x

Hazard Rate

We define the hazard rate for a distribution function F with density f to be

$$\lambda(x) = \frac{f(x)}{1 - F(x)}$$

which we can use to uniquely identify a distribution

$$\begin{aligned}
\int_0^x \lambda(t)\, dt &= \int_0^t \frac{f(x)}{1-F(x)} dt = \int_0^t \frac{\frac{d}{dt}F(x)}{1-F(x)} dt \\
&= -\log(1 - F(t)) + \log(1 - F(0)) \\
\int_0^x \lambda(t)\, dt &= -\log(1 - F(t)) \\
1 - F(t) &= \exp\left(-\int_0^x \lambda(t)\, dt\right) \\
F(t) &= 1 - \exp\left(-\int_0^x \lambda(t)\, dt\right)
\end{aligned}$$

Some Quick Definitions

Monotonically increasing (increasing, non-decreasing) function:

$$x \leq y \implies f(x) \leq f(y)$$

Monotonically decreasing (decreasing, non-increasing) function:

$$x \leq y \implies f(x) \geq f(y)$$

Strictly increasing function:

$$x < y \implies f(x) < f(y)$$

Strictly decreasing function:

$$x < y \implies f(x) > f(y)$$

Change of Variables for Continuous RV

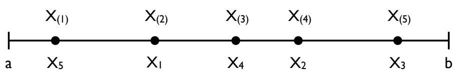
Let X be a random variable with density $f_X(x)$ on the range (a, b) and let $Y = g(X)$ which has the range $(g(a), g(b))$, if $g(x)$ is either strictly increasing or decreasing on (a, b) then

$$f_Y(y) = f_X(x) \bigg/ \left|\frac{dy}{dx}\right|$$

Order Statistics

Let X_1, X_2, X_3, X_4, X_5 be iid random variables with a distribution F with a range of (a, b). We can relabel these X's such that their labels correspond to arranging them in increasing order so that

$$X_{(1)} \le X_{(2)} \le X_{(3)} \le X_{(4)} \le X_{(5)}$$



In the case where the distribution F is continuous we can make the stronger statement that

$$X_{(1)} < X_{(2)} < X_{(3)} < X_{(4)} < X_{(5)}$$

Since $P(X_i = X_j) = 0$ for all $i \neq j$ for continuous random variables.

Order Statistics, cont.

For $X_1, X_2, \ldots, X_n$ iid random variables X_k is the kth smallest X, usually called the kth order statistic.

$X_{(1)}$ is therefore the smallest X and

$$X_{(1)} = \min(X_1, \ldots, X_n)$$

Similarly, $X_{(n)}$ is the largest X and

$$X_{(n)} = \max(X_1, \ldots, X_n)$$

Distributions of order statistics

For $X_1, X_2, \ldots, X_n$ iid continuous random variables with pdf f and cdf F then

$$f_{(1)}(x) = nf(x)(1 - F(x))^{n-1}$$
$$f_{(k)}(x) = nf(x)\binom{n-1}{k-1}F(x)^{k-1}(1 - F(x))^{n-k}$$
$$f_{(n)}(x) = nf(x)F(x)^{n-1}$$

$$F_{(1)}(x) = 1 - (1 - F(x))^n$$
$$F_{(n)}(x) = F(x)^n$$

Uniform Distribution

If X is a random variable with constant density on (a, b) then X is said to be Uniformly distributed on (a, b), $X \sim \text{Unif}(a, b)$, then

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

$$E(X) = \frac{b+a}{2}$$
$$Var(X) = \frac{(b-a)^2}{12}$$

Normal Distribution

If X is random variable with a normal distribution with a mean μ and variance σ^2, $X \sim \mathcal{N}(\mu, \sigma^2)$, then

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$
$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$E(X) = \mu$$
$$Var(X) = \sigma^2$$
$$\text{Mode}(X) = \mu$$

Exponential Distribution

Let X be a random variable that reflects the time between events which occur continuously with a given rate λ, $X \sim \text{Exp}(\lambda)$

$$f(x|\lambda) = \lambda e^{-\lambda x}$$
$$P(X \le x) = F(x|\lambda) = 1 - e^{-\lambda x}$$

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1} = \left(\frac{\lambda}{\lambda - t}\right)$$

$$E(X) = \lambda^{-1} \qquad E(X^n) = \frac{n!}{\lambda^n}$$
$$Var(X) = \lambda^{-2} \qquad Median(X) = \frac{\log 2}{\lambda}$$

Memoryless property - $P(X > s + t | X > s) = P(X > t)$
Minimum of Exponentials - $\min(X_1, \ldots, X_n) \sim \text{Exp}(\lambda_1 + \cdots + \lambda_n)$

Gamma Function

Based on the relationship for the nth raw moment, μ'_n, of an exponential distribution

$$E(X^n) = \frac{n!}{\lambda^n}$$

Let set $\lambda = 1$ and define an new value $\alpha = n + 1$

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

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

Using a tradition definition of the factorial it only makes sense when $n \in \mathbb{N}$ but we can use this new definition of the Gamma function $\Gamma(\alpha)$ for any $\alpha \in \mathbb{R}^+$.

Commonly used with the Gamma, Beta, and negative binomial distribution to generalize a parameter to $\mathbb{R}^+$.

Erlang Distribution

Let X reflect the time until the nth event occurs when the events occur according to a Poisson process with rate λ, $X \sim \text{Er}(n, \lambda)$

$$f(x|n,\lambda) = \frac{e^{-\lambda x}\lambda^n x^{n-1}}{(n-1)!}$$
$$F(x|n,\lambda) = \sum_{j=n}^{\infty} \frac{e^{-\lambda x}(\lambda x)^j}{j!}$$

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^n$$

$$E(X) = n/\lambda$$
$$Var(X) = n/\lambda^2$$

Gamma Distribution

We can generalize the Erlang distribution by using the gamma function instead of the factorial function, thereby allowing for $\mathbb{R}^+$ values of n. Often the distribution is reparameterized such that $\theta = 1/\lambda$, $X \sim \text{Gamma}(n, \theta)$.

$$f(x|n,\theta) = \frac{e^{-x/\theta} x^{n-1}}{\theta^n \Gamma(n)}$$

$$F(x|n,\theta) = \frac{\int_0^x e^{-t/\theta} t^{n-1}\, dt}{\theta^n \Gamma(n)} = \frac{\gamma(n, x/\theta)}{\Gamma(n)}$$

$$M_X(t) = \left(\frac{1}{1-\theta t}\right)^n$$

$$E(X) = n\theta$$

$$Var(X) = n\theta^2$$

Beta Distribution

If $X \sim \text{Beta}(r, s)$ then

$$f(x) = \frac{1}{B(r,s)} x^{r-1}(1-x)^{s-1}$$

$$F(x) = \int_0^x \frac{1}{B(r,s)} x^{r-1}(1-x)^{s-1} dx = \frac{B_x(r,s)}{B(r,s)}$$

$$B(r,s) = \int_0^1 x^{r-1}(1-x)^{s-1} dx = \frac{(r-1)!(s-1)!}{(r+s-1)!} = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}$$

$$B_x(r,s) = \int_0^x x^{r-1}(1-x)^{s-1} dx$$

$$E(X) = \frac{r}{r+s}$$

$$Var(X) = \frac{rs}{(r+s)^2(r+s+1)}$$

Joint Distributions - Example

Draw two socks at random, without replacement, from a drawer full of twelve colored socks:

6 black, 4 white, 2 purple

Let B be the number of Black socks, W the number of White socks drawn, then the distributions of B and W are given by:

	0	1	2
P(B=k)	$\frac{6}{12}\frac{5}{11} = \frac{15}{66}$	$2\frac{6}{12}\frac{6}{11} = \frac{36}{66}$	$\frac{6}{12}\frac{5}{11} = \frac{15}{66}$
P(W=k)	$\frac{8}{12}\frac{7}{11} = \frac{28}{66}$	$2\frac{4}{12}\frac{8}{11} = \frac{32}{66}$	$\frac{4}{12}\frac{3}{11} = \frac{6}{66}$

Note - $B \sim \text{HyperGeo}(12, 6, 2) = \frac{\binom{6}{k}\binom{6}{2-k}}{\binom{12}{2}}$ and $W \sim \text{HyperGeo}(12, 4, 2) = \frac{\binom{4}{k}\binom{8}{2-k}}{\binom{12}{2}}$

Joint Distributions - Example, cont.

Let B be the number of Black socks, W the number of White socks drawn, then the distributions of B and W are given by:

		W: 0	1	2	
B	0	$\frac{1}{66}$	$\frac{8}{66}$	$\frac{6}{66}$	$\frac{15}{66}$
	1	$\frac{12}{66}$	$\frac{24}{66}$	0	$\frac{36}{66}$
	2	$\frac{15}{66}$	0	0	$\frac{15}{66}$
		$\frac{28}{66}$	$\frac{32}{66}$	$\frac{6}{66}$	$\frac{66}{66}$

$$P(B=b, W=w) = \frac{\binom{6}{b}\binom{4}{w}\binom{2}{2-b-w}}{\binom{12}{2}}$$

Marginal Distribution

Note that the column and row sums are the distributions of B and W respectively.

$$P(B=b) = P(B=b, W=0) + P(B=b, W=1) + P(B=b, W=2)$$

$$P(W=w) = P(B=0, W=w) + P(B=1, W=w) + P(B=2, W=w)$$

These are the *marginal* distributions of B and W. In general,

$$P(X=x) = \sum_y P(X=x, Y=y) = \sum_y P(X=x|Y=y)P(Y=y)$$

Conditional Distribution

Conditional distributions are defined as we have seen previously with

$$P(X=x|Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{\text{joint pmf}}{\text{marginal pmf}}$$

Therefore the pmf for white socks given no black socks were drawn is

$$P(W=w|B=0) = \frac{P(W=w, B=0)}{P(B=0)} = \begin{cases} \frac{1}{66} \Big/ \frac{15}{66} = \frac{1}{15} & \text{if } W=0 \\ \frac{8}{66} \Big/ \frac{15}{66} = \frac{8}{15} & \text{if } W=1 \\ \frac{6}{66} \Big/ \frac{15}{66} = \frac{6}{15} & \text{if } W=2 \end{cases}$$

Expectation of Joint Distributions

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) P(X=x, Y=y)$$

For example we can define $g(x,y) = x \cdot y$ then

$$\begin{aligned} E(BW) =& (0 \cdot 0 \cdot 1/66) + (0 \cdot 1 \cdot 8/66) + (0 \cdot 2 \cdot 6/66) \\ &+ (1 \cdot 0 \cdot 12/66) + (1 \cdot 1 \cdot 24/66) + (1 \cdot 2 \cdot 0/66) \\ &+ (2 \cdot 0 \cdot 15/66) + (2 \cdot 1 \cdot 0/66) + (1 \cdot 2 \cdot 0/66) \\ =& 24/66 = 4/11 \end{aligned}$$

Note that $E(BW) \neq E(B)E(W)$ since

$$\begin{aligned} E(B)E(W) =& (0 \cdot 15/66 + 1 \cdot 36/66 + 2 \cdot 15/66) \\ &\times (0 \cdot 28/66 + 1 \cdot 32/66 + 2 \cdot 6/66) \\ =& 66/66 \times 44/66 = 2/3 \end{aligned}$$

This implies that B and W are not independent.

Independence

Remember that $Cov(X,Y) = 0$ when X and Y are independent.

$$\begin{aligned} Cov(B,W) &= E[(B - E[B])(W - E[W])] \\ &= E(BW) - E(B)E(W) \\ &= 4/11 - 2/3 = -10/33 = -0.30303 \end{aligned}$$

Expectation of Conditional Probability

Works like any other distribution

$$E(X|Y=y) = \sum_x xP(X=x|Y=y)$$

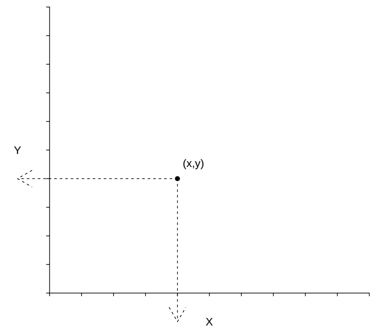
Therefore we can calculating things like conditional mean and variance,

$$E(W|B=0) = 0 \cdot 1/15 + 1 \cdot 8/15 + 2 \cdot 6/15 = 20/15 = 1.333$$
$$E(W^2|B=0) = 0^2 \cdot 1/15 + 1^2 \cdot 8/15 + 2^2 \cdot 6/15 = 32/15 = 2.1333$$

$$Var(W|B=0) = E(W^2|B=0) - E(W|B=0)^2$$
$$= 32/15 - (4/3)^2 = 16/45 = 0.3556$$

Joint CDF

$$F(x,y) = P[X \le x, Y \le y]$$
$$= P[(X,Y) \text{ lies south-west of the point } (x,y)]$$



Joint CDF, cont.

The joint Cumulative distribution function follows the same rules as the univariate CDF,

Univariate definition:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(z)dz$$

$$\lim_{x \to -\infty} F(x) = 0 \qquad \lim_{x \to \infty} F(x) = 1 \qquad x \le y \Rightarrow F(x) \le F(y)$$

Bivariate definition:

$$F(x,y) = P(X \le x, Y \le y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x,y)\; dx\; dy$$

$$\lim_{x,y \to -\infty} F(x,y) = 0 \qquad \lim_{x,y \to \infty} F(x,y) = 1 \qquad x \le x', y \le y' \Rightarrow F(x,y) \le F(x',y')$$

Marginal Distributions

We can define marginal CDFs using the joint CDF by setting one of the values to infinity:

$$F(x,\infty) = P(X \le x, Y \le \infty) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(x,y)\; dy\; dy$$
$$= P(X \le x) = F_X(x)$$

$$F(\infty,y) = P(X \le \infty, Y \le y) = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f(x,y)\; dx\; dy$$
$$= P(Y \le y) = F_Y(y)$$

Joint pdf

Similar to the CDF the probability density function follows the same general rules in two dimensions,

Univariate definition:

$f(x) \geq 0$ for all x $\qquad f(x) = \frac{d}{dx}F(x) \qquad \int_{-\infty}^{\infty} f(x)dx = 1$

Bivariate definition:

$$f(x,y) \geq 0 \text{ for all } (x,y)$$

$$f(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}F(x,y)$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\; dx\; dy = 1$$

Marginal pdfs

Marginal pdfs are derived by integrating out one of the random variables.

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\; dy$$

$$f_Y(x) = \int_{-\infty}^{\infty} f(x,y)\; dx$$

Previously we defined independence in terms of, X and Y are independent if and only if $E(XY) = E(X)E(Y)$.

An equivalent definitions is, X and Y are independent if and only if $f(x,y) = f_X(x)f_Y(y)$.

Probability and Expectation

Univariate definition:

$$P(X \in A) = \int_A f(x)\; dx$$

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

Bivariate definition:

$$P(X \in A, Y \in B) = \int_A \int_B f(x,y)\; dx\; dy$$

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