13 Lecture 13: Feb 17

Last time

- Probability review

Today

- HW1 review
- Probability review, cont

Reference:

- Statistical Inference, 2nd Edition, by George Casella & Roger L. Berger
- Review of Probability Theory by Arian Maleki and Tom Do

Binomial mean

IF X has binomial distribution, i.e. $X \sim binomial(n, p)$, its pmf is given by

$$\Pr(X = x) = \left(\begin{array}{c} \text{n} \\ \text{x} \end{array} \right) p^x (1-p)^{n-x}, \quad x = 0, 1, \ldots, n,$$

where n is a positive integer, $0 \leqslant p \leqslant 1$, and for every fixed pair n and p the pmf sums to 1. The expected value of a binomial random variable is then given by

$$\mathbf{E}(X) = \sum_{x=0}^{n} x \left(\begin{array}{c} \text{n} \\ \text{x} \end{array} \right) p^x (1-p)^{n-x}$$

Now, use the identity $x \left(\begin{array}{c} \text{n} \\ \text{x} \end{array} \right) = n \left(\begin{array}{c} \text{n - 1} \\ \text{x - 1} \end{array} \right)$ to derive the Expected value.

properties:

Let X be a random variable and let a, b and c be constants. Then for any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,

1. $\mathbf{E}(a \cdot g_1(X) + b \cdot g_2(X) + c) = a\mathbf{E}(g_1(X)) + b\mathbf{E}(g_2(X)) + c.$
2. If $g_1(x) \geqslant 0$ for all x, then $\mathbf{E}(g_1(X)) \geqslant 0$.
3. If $g_1(x) \geqslant g_2(x)$ for all x, then $\mathbf{E}(g_1(X)) \geqslant \mathbf{E}(g_2(X))$.
4. If $a \leqslant g_1(x) \leqslant b$ for all x, then $a \leqslant \mathbf{E}(g_1(X)) \leqslant b$.

Moments

The various moments of a distribution are an important class of expectations.

Definition: For each integer n, the n^{th} moment of X (or $F_X(x)$), μ'_n, is

$$\mu'_n = \mathbf{E}\left(X^n\right).$$

The n^{th} central moment of X, μ_n, is

$$\mu_n = \mathbf{E}\left((X-\mu)^n\right),$$

where $\mu = \mu'_1 = \mathbf{E}\left(X\right)$.

Variance

Definition: The variance of a random variable X is its second central moment, $\mathbf{Var}\left(X\right) = \mathbf{E}\left((X - EX)^2\right)$. The positive square root of $\mathbf{Var}\left(X\right)$ is the standard deviation of X.

Exponential variance

Let X have the exponential(λ) distribution, $X \sim Exp(\lambda)$. Then the variance of X is

properties

1. $\mathbf{Var}\left(aX+b\right) = a^2\mathbf{Var}\left(X\right)$.
   proof:

2. $\mathbf{Var}\left(X\right) = \mathbf{E}\left(X^2\right) - \left(\mathbf{E}\left(X\right)\right)^2$.
   proof:

Moment generating function

Definition: Let X be a random variable with cdf F_X. The moment generating function or mgf of X (or F_X), denoted by $M_X(t)$, is

$$M_X(t) = \mathbf{E}\left(e^{tX}\right),$$

provided that the expectation exists for t in some neighborhood of 0. That is, there exists an $h > 0$ such that for all t in $-h < t < h$, $\mathbf{E}\left(e^{tX}\right)$ exists. If the expectation does not exist in a neighborhood of 0, we say that the moment generating function does not exist.

Property: If X has mgf $M_X(t)$, then

$$\mathbf{E}\left(X^n\right) = M_X^{(n)}(0),$$

where we define

$$M_X^{(n)}(0) = \left.\frac{d^n}{dt^n} M_X(t)\right|_{t=0}.$$

Some common random variables

Discrete random variables

- $X \sim Bernoulli(p)$ (where $0 \leqslant p \leqslant 1$):

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

- $X \sim Binomial(n,p)$ (where $0 \leqslant p \leqslant 1$):

$$\Pr(x) = \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x}$$

- $X \sim Geometric(p)$ (where $0 \leqslant p \leqslant 1$):

$$\Pr(x) = p(1-p)^{x-1}$$

- $X \sim Poisson(\lambda)$ (where $\lambda > 0$):

$$\Pr(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

Continuous random variables

- $X \sim Uniform(a,b)$ (where $a < b$):

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

- $X \sim Exponential(\lambda)$ (where $\lambda > 0$):

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

- $X \sim Normal(\mu, \sigma^2)$:

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

The following table provides a summary of some of the properties of these distributions.

Distribution	PDF or PMF	Mean	Variance
$Bernoulli(p)$	$\begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$	p	$p(1-p)$
$Binomial(n,p)$	$\begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{n-x}$, for $0 \leqslant k \leqslant n$	np	$np(1-p)$
$Geometric(p)$	$p(1-p)^{x-1}$, for $k = 1, 2, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
$Poisson(\lambda)$	$e^{-\lambda}\frac{\lambda^x}{x!}$, for $k = 1, 2, \ldots$	λ	λ
$Uniform(a,b)$	$\frac{1}{b-a} I(a \leqslant x \leqslant b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$Gaussian(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$	μ	σ^2
$Exponential(\lambda)$	$\lambda e^{-\lambda x} I(x \geqslant 0)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Chi-square, t-, and F-Distributions

Let $Z_1, Z_2, \ldots, Z_k \overset{iid}{\sim} N(0,1)$, then $X^2 \equiv Z_1^2 + Z_2^2 + \cdots + Z_k^2 \sim \chi_k^2$ (with k degrees of freedom). If $X \sim \chi_k^2$

$$\mathbf{E}(X) = k$$
$$\mathbf{Var}(X) = 2k.$$

Student's t versus χ^2

If $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, then

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

When σ is unknown,

$$\frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}, \quad \text{where } \hat{\sigma} = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}}.$$

Note that

$$\begin{aligned} \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} &= \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \cdot \frac{1}{\frac{\hat{\sigma}}{\sigma}} \\ &= Z \cdot \frac{1}{\sqrt{\frac{\sum(X_i - \bar{X})^2}{(n-1)\sigma^2}}} \\ &= \frac{Z}{\sqrt{\frac{\chi_{n-1}^2}{n-1}}} \end{aligned}$$

F versus χ^2

$$F_{ndf,ddf} \equiv \frac{\chi_{ndf}^2/ndf}{\chi_{ddf}^2/ddf}$$

t versus χ^2

$$\begin{aligned} t_k &= \frac{Z}{\sqrt{\chi_k^2/k}} \\ &= \frac{\sqrt{\chi_1^2/1}}{\sqrt{\chi_k^2/k}} \\ &= \sqrt{F_{1,k}} \end{aligned}$$

or, in other words, $t_k^2 = F_{1,k}$

Random vectors and matrices

The cdf for random vector

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \text{ is } F_{\mathbf{Y}}(\mathbf{y}) = \Pr(Y_1 \leqslant y_1, Y_2 \leqslant y_2, \ldots, Y_n \leqslant y_n)$$

If a joint pdf exists, then $f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Y}}(y_1, \ldots, y_n)$ and

$$F_{\mathbf{Y}}(\mathbf{y}) = \int\limits_{-\infty}^{y_1} \int\limits_{-\infty}^{y_2} \cdots \int\limits_{-\infty}^{y_n} f_{\mathbf{Y}}(\mathbf{t}) d\mathbf{t}$$

Moments

$$\begin{aligned}
\mathbf{E}(\mathbf{Y}) = \mu_{\mathbf{Y}} &= \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \\
\mathbf{Var}(\mathbf{Y}) &= \mathbf{E}\left((\mathbf{Y} - \mu_{\mathbf{Y}})(\mathbf{Y} - \mu_{\mathbf{Y}})^T\right) \\
&= \mathbf{E}\left(\begin{bmatrix} (Y_1 - \mu_1)^2 & (Y_1 - \mu_1)(Y_2 - \mu_2) & \ldots \\ (Y_2 - \mu_2)(Y_1 - \mu_1) & (Y_2 - \mu_2)^2 & \ldots \\ \ldots & & \end{bmatrix}\right) \\
&= \mathbf{E}\left([(Y_i - \mu_i)(Y_j - \mu_j), i = 1, 2, \ldots, n, j = 1, 2, \ldots, n]\right) \\
&= (\sigma_{ij})_{i=1,2,\ldots,n; j=1,2,\ldots,n}
\end{aligned}$$

where $\sigma_{ij} = Cov(Y_i, Y_j)$

Linear functions

Let $\mathbf{X} \in \mathbb{R}^{k\times 1}, \mathbf{Y} \in \mathbb{R}^{n\times 1}$ and $\mathbf{A} \in \mathbb{R}^{k\times 1}$, $\mathbf{B} \in \mathbb{R}^{k\times n}$ be non-random, then

$$\begin{aligned}
\underset{k\times 1}{\mathbf{X}} &= \underset{k\times 1}{\mathbf{A}} + \underset{k\times n}{\mathbf{B}}\,\underset{n\times 1}{\mathbf{Y}} \\
\mathbf{E}(\mathbf{X}) &= \mathbf{A} + \mathbf{B}\mathbf{E}(\mathbf{Y}) \\
\mathbf{Var}(\mathbf{X}) &= \mathbf{B}\mathbf{Var}(\mathbf{Y})\mathbf{B}^T
\end{aligned}$$

Sums of random vectors

$$\begin{aligned}
\underset{n\times 1}{\mathbf{X}} &= \underset{n\times 1}{\mathbf{Y}} + \underset{n\times 1}{\mathbf{Z}} \\
\mathbf{E}(\mathbf{X}) &= \mathbf{E}(\mathbf{Y}) + \mathbf{E}(\mathbf{Z}) = \mathbf{E}(\mathbf{Y} + \mathbf{Z})
\end{aligned}$$

Note that there is no independence assumed above.

$$\mathbf{Var}(\mathbf{X}) = \mathbf{Var}(\mathbf{Y} + \mathbf{Z}) = \mathbf{Var}(\mathbf{Y}) + \mathbf{Var}(\mathbf{Z}) + Cov(\mathbf{Y}, \mathbf{Z}) + Cov(\mathbf{Z}, \mathbf{Y})$$

If $\mathbf{Y}, \mathbf{Z}$ are uncorrelated, then $\mathbf{Var}(\mathbf{X}) = \mathbf{Var}(\mathbf{Y}) + \mathbf{Var}(\mathbf{Z})$