BIOS 660/BIOS 672 (3 Credits):
Probability and Statistical Inference I

Jianwen Cai

https://sakai.unc.edu/portal/site/bios660-bios672-3-credits

Notes 8

Expectation

- The *expected value* or *mean* of a rv X, denoted E(X) or EX is

$$\text{E}X = \begin{cases} \int_{-\infty}^{\infty} x\, f_X(x)\, dx, & X \text{ continuous} \\ \sum_{x \in \mathcal{X}} x\, f_X(x), & X \text{ discrete} \end{cases}$$

  Provided the integral or summation exists.

- This is generalized for a function of a random variable $g(X)$ as

$$\text{E}g(X) = \begin{cases} \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx, & X \text{ continuous} \\ \sum_{x \in \mathcal{X}} g(x)\, f_X(x), & X \text{ discrete} \end{cases}$$

  Notice we could also find the pdf or pmf of Y and use the first definition. Both give the same answer (HW).

Expectation of some continuous variables

- Let $X \sim$ U$[a, b]$. Then

$$\text{E}X = \int_a^b x \frac{1}{b-a}\, dx = \frac{a+b}{2}$$

- Let $X \sim$ Exp(1), $f_X(x) = e^{-x}\, 1(x > 0)$. Then

$$\text{E}X = \int_0^\infty x e^{-x}\, dx = -xe^{-x}\Big|_0^\infty + \int_0^\infty e^{-x}\, dx = 1$$

- Let $X \sim$ N$(0, 1)$. Then

$$\text{E}X = \int_{-\infty}^\infty x \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = 0$$

  (since the above integral is an odd function)

Expectation of some discrete variables

- Let $X = 1(A)$, where A is a Borel set. Then

$$\mathsf{E}X = 0 \cdot P(A^c) + 1 \cdot P(A) = P(A)$$

- Let $X \sim \text{Binomial}(n, p)$ for n positive integer and $0 < p < 1$ (n is the number of independent identical binary trials and p is the probability of success). Then

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, \ldots, n$$

and

$$\begin{aligned} \mathsf{E}X &= \sum_{x=0}^{n} x \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x} \\ &= np \sum_{x=1}^{n} \frac{(n-1)!}{(n-x)!(x-1)!} p^{x-1} (1-p)^{n-x} \\ (\text{let } y = x-1) \quad &= np \sum_{y=0}^{n-1} \frac{(n-1)!}{(n-1-y)!y!} p^y (1-p)^{n-1-y} \\ &= np \end{aligned}$$

An easier way later.

cont.

- Let $X \sim \text{Geom}(p)$ (e.x. toss till first success), $f_X(x) = pq^{x-1}$, $x = 1, 2, \ldots$, where $q = 1 - p$. Then

$$\begin{aligned} \mathsf{E}X &= \sum_{x=1}^{\infty} x \cdot pq^{x-1} = \sum_{x=1}^{\infty} p \frac{d}{dq}(q^x) \\ &= p \frac{d}{dq} \sum_{x=1}^{\infty} q^x = p \frac{d}{dq}\left(\frac{1}{1-q} - 1\right) \\ &= p \frac{1}{(1-q)^2} = \frac{1}{p} \end{aligned}$$

Note: The differentiation operator can be moved outside the summation sign because the geometric series converges uniformly.

Mean = area under survival curve

For a non-negative random variable X (i.e. $f(x) = 0$ for $x < 0$),

$$\mathsf{E}(X) = \begin{cases} \int_0^\infty (1 - F(x))dx, & X \text{ continuous} \\ \sum_{x=0}^\infty (1 - F(x)), & X \text{ discrete} \end{cases}$$

Proof: Homework! (Exercise 2.14)

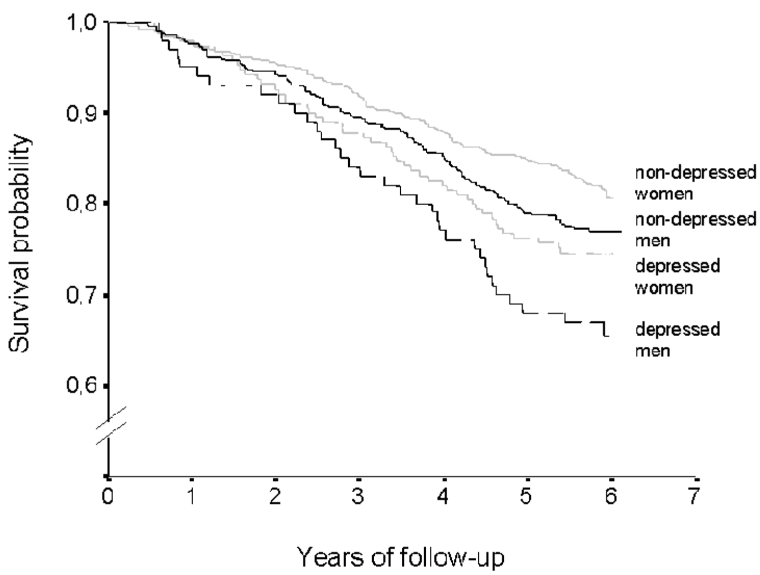
Example: Let $X \sim \mathsf{Exp}(\lambda)$ with

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

Then

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

Survival curves



Properties of expectation

$$\begin{aligned}
\mathsf{E}[ag(X)+c] \quad &= \quad a\mathsf{E}[g(X)]+c, \qquad a,\, c \text{ constants} \\
\mathsf{E}[g_1(X)+g_2(X)] \quad &= \quad \mathsf{E}[g_1(X)]+\mathsf{E}[g_2(X)] \\
\text{If } g_1(x) \geq 0\ \forall x, \quad &\text{then} \quad \mathsf{E}g_1(X) \geq 0 \\
\text{If } a \leq g(X) \leq b\ \forall x, \quad &\text{then} \quad a \leq \mathsf{E}g(X) \leq b
\end{aligned}$$

(Proofs are immediate page 57)

Note: The linearity property above applies to two random variables X and Y that have the same distribution:

$$\mathsf{E}[g_1(X)+g_2(Y)] = \mathsf{E}[g_1(X)]+\mathsf{E}[g_2(Y)]$$

We will see later that this is true even if X and Y do not have the same distribution.

Method of indicators

An example of how the above properties are useful.

Let $X \sim$ Binomial(n,p) for n positive integer and $0 < p < 1$ (n is the number of independent identical binary trials and p is the probability of success). We can write

$$X = \sum_{i=1}^{n} I_i$$

where I_i is the indicator that the i^{th} trial is a success. We have

$$\mathsf{E}I_i = p$$

Therefore

$$\mathsf{E}X = \sum_{i=1}^{n} \mathsf{E}I_i = \sum_{i=1}^{n} p = np$$

Minimal property of the mean

Assuming all integrals exist,

$$\min_b \mathsf{E}(X-b)^2 = \mathsf{E}(X-\mathsf{E}X)^2 = \mathsf{Var}X$$

Two proofs:

- By differentiation with respect to b (homework). Requires more assumptions.
- By sum-of-squares decomposition.

Moments

For a random variable X, the expectation of the polynomials $g(X) = X^r$, $r = 0, 1, 2, \ldots$ are called the *moments* of X:

$$m_r = \mathsf{E}(X^r), \qquad r = 0, 1, 2, \ldots$$

These are sometimes called *non-central* moments or *moments about the origin*.

Notes:

- $m_0 = 1$.
- m_1 is the *mean*, usually denoted by $m_1 = \mu$.

Central moments

The r^{th} *central moment* of X is

$$\mu_r = \mathsf{E}[X - \mathsf{E}X]^r = \mathsf{E}[X - \mu]^r, \qquad r = 0, 1, 2, \ldots$$

These are sometimes called *moments about the mean*.

Notes:

- $\mu_0 = 1$.
- $\mu_1 = 0$.
- μ_2 is the *variance*.
- μ_3 is related to the *skewness*.
- μ_4 is related to the *kurtosis*.

Variance

The second central moment μ_2 is called the *variance of* X and is usually denoted by σ^2:

$$\sigma^2 = \mu_2 = \mathsf{Var}(X) = \mathsf{E}[X - \mu]^2 = \mathsf{E}(X^2) - \mu^2$$

Notes:

- Useful property: For $a, b \in \mathbb{R}$,
$$\mathsf{Var}(aX + b) = a^2 \mathsf{Var}(X)$$
- Notice that $\mathsf{E}(X^2) \neq [\mathsf{E}(X)]^2$. In fact, $\mathsf{E}(X^2) \geq [\mathsf{E}(X)]^2$ because
$$\mathsf{Var}(X) = \mathsf{E}(X^2) - [\mathsf{E}(X)]^2 \geq 0$$

Skewness

The *skewness* of a rv X is defined as

$$\alpha_3 = \frac{\mu_3}{(\mu_2)^{3/2}}, \qquad \mu_3 = \mathsf{E}[(X-\mu)^3]$$

Skewness measures the symmetry of a distribution.

$$\begin{aligned} \mu_3 = 0 &\quad \Rightarrow \quad \text{symmetric} \\ \mu_3 \geq 0 &\quad \Rightarrow \quad \text{right skew} \\ \mu_3 \leq 0 &\quad \Rightarrow \quad \text{left skew} \end{aligned}$$

Central and non-central moments

Central moments can be written as a function of non-central moments and vice-versa:

$$\begin{aligned} \mu_2 &= m_2 - m_1^2 \\ \mu_3 &= m_3 - 3m_1m_2 + 2m_1^3 \end{aligned}$$

etc.

Method of indicators (again)

Recall that $X \sim$ Binomial(n,p) can be written as $X = \sum_{i=1}^n I_i$ where I_i is the indicator that the i^{th} trial out of n is a success. We have

$$\mathsf{E}I_i = p, \qquad \mathsf{E}I_i^2 = p, \qquad \mathsf{Var}I_i = p(1-p)$$

Therefore

$$\begin{aligned}\mathsf{E}X^2 &= \mathsf{E}\left(\sum_{i=1}^n I_i\right)^2 = \mathsf{E}\sum_{i=1}^n\sum_{j=1}^n I_i I_j \\ &= \mathsf{E}\left(\sum_{i=1}^n I_i + \sum_{i\neq j} I_i I_j\right) = \sum_{i=1}^n p + \sum_{i\neq j} p^2 = np + (n^2-n)p^2\end{aligned}$$

and

$$\mathsf{Var}X = \mathsf{E}X^2 - \mathsf{E}^2X = np + (n^2-n)p^2 - (np)^2 = np(1-p)$$