Lecture 03: Transformations and Expectations

Mathematical Statistics I, MATH 60061/70061

Tuesday September 7, 2021

Reference: Casella & Berger, 2.1-2.2

Distributions of functions of a random variable

If $X \in \mathcal{X}$ is a random variable with CDF $F_X(x)$, then any function of X, say $g(X)$ is also a random variable.

Let $Y = g(X) \in \mathcal{Y}$. We can express the distribution of Y in terms of the functions F_X and g.

For any set $A \subset \mathcal{Y}$,

$$\begin{aligned} P(Y \in A) &= P(g(X) \in A) \\ &= P(\{x \in \mathcal{X} : g(x) \in A\}) \\ &= P(X \in g^{-1}(A)). \end{aligned}$$

Distributions of functions of a discrete random variable

If X is a discrete random variable, then $\mathcal{X}$ is countable. The sample space for $Y = g(X)$ is $\mathcal{Y} = \{y : y = g(x), x \in \mathcal{X}\}$, which is also a countable set.

Thus, Y is also a discrete random variable, and the PMF for Y is

$$f_Y(y) = P(Y = y) = \sum_{x \in g^{-1}(y)} P(X = x) = \sum_{x \in g^{-1}(y)} f_X(x),$$

for $y \in \mathcal{Y}$, and $f_Y(y) = 0$ for $y \notin \mathcal{Y}$.

Binomial transformation

A discrete random variable has a **Binomial distribution** if its PMF is of the form

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

where n is a positive integer and $0 \leq p \leq 1$.

Values such as n and p that can be set to different values, producing different probability distributions, are called **parameters**.

Binomial transformation

Consider the random variable $Y = g(X) = n - X$. The sample spaces for X and Y are $\mathcal{X} = \{0, 1, \ldots, n\}$ and $\mathcal{Y} = \{y : y = g(x), x \in \mathcal{X}\} = \{0, 1, \ldots, n\}$, respectively.

For any $y \in \mathcal{Y}$,

$$
\begin{aligned}
f_Y(y) &= \sum_{x \in g^{-1}(y)} f_X(x) \\
&= f_X(n - y) \\
&= \binom{n}{n-y} p^{n-y} (1-p)^{n-(n-y)} \\
&= \binom{n}{y} (1-p)^y p^{n-y}.
\end{aligned}
$$

Thus, Y also has a binomial distribution, but with parameters n and $1 - p$.

Distributions of functions of continuous random variables

If X and $Y = g(X)$ are continuous random variables, the CDF of Y is

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(\{x \in \mathcal{X} : g(x) \leq y\}) \\ &= \int_{\{x \in \mathcal{X} : g(x) \leq y\}} f_X(x) dx. \end{aligned}$$

Sometimes it may be difficult to identify $\{x \in \mathcal{X} : g(x) \leq y\}$ and carry out the integration of $f_X(x)$ over this region.

Transformation with monotone functions

A function $g(x)$ is monotone if it is either

$$u > v \Rightarrow g(u) > g(v) \quad \text{(increasing)}$$

or

$$u < v \Rightarrow g(u) > g(v) \quad \text{(decreasing)}.$$

Let $y = g(x)$. If g is monotone, then g^{-1} is single-valued

- If g is increasing

$$\begin{aligned}\{x \in \mathcal{X} : g(x) \leq y\} &= \{x \in \mathcal{X} : g^{-1}(g(x)) \leq g^{-1}(y)\} \\ &= \{x \in \mathcal{X} : x \leq g^{-1}(y)\}.\end{aligned}$$

- If g is decreasing

$$\begin{aligned}\{x \in \mathcal{X} : g(x) \leq y\} &= \{x \in \mathcal{X} : g^{-1}(g(x)) \geq g^{-1}(y)\} \\ &= \{x \in \mathcal{X} : x \geq g^{-1}(y)\}.\end{aligned}$$

CDFs of monotone functions

- If g is increasing

$$\begin{aligned}
F_Y(y) &= \int_{\{x\in\mathcal{X}: x\le g^{-1}(y)\}} f_X(x)dx \\
&= \int_{-\infty}^{g^{-1}(y)} f_X(x)dx \\
&= F_X(g^{-1}(y)).
\end{aligned}$$

- If g is decreasing

$$\begin{aligned}
F_Y(y) &= \int_{g^{-1}(y)}^{\infty} f_X(x)dx \\
&= 1 - F_X(g^{-1}(y)).
\end{aligned}$$

PDFs of monotone functions

Let X have PDF $f_X(x)$ and let $Y = g(X)$, where g is a monotone function. Suppose that $f_X(x)$ is continuous on $\mathcal{X}$ and that $g^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$. Then, by chain rule, the PDF of Y is given by

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy} F_Y(y) \\
&= \begin{cases} f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) & \text{if } g \text{ is increasing} \\ -f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) & \text{if } g \text{ is decreasing} \end{cases} \\
&= f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|
\end{aligned}
$$

for $y \in \mathcal{Y}$, and $f_Y(y) = 0$ otherwise.

Piecewise transformation

Let X have pdf $f_X(x)$, let $Y = g(X)$. Suppose there exists a partition, $A_0, A_1, \ldots, A_k$, of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose there exist functions $g_1(x), \ldots, g_k(x)$, defined on $A_1, \ldots, A_k$, respectively, satisfying:

1. $g(x) = g_i(x)$, for $x \in A_i$,
2. $g_i(x)$ is monotone on A_i,
3. the set $\mathcal{Y} = \{y : y = g_i(x) \text{ for some } x \in A_i\}$ is the same for each $i = 1, \ldots, k$, and
4. $g_i^{-1}(x)$ has a continuous derivative on $\mathcal{Y}$, for each $i = 1, \ldots, k$.

Then

$$f_Y(y) = \begin{cases} \sum_{i=1}^k f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & \text{otherwise.} \end{cases}$$

Chi-Squared PDF

Let X have the standard **Normal distribution**,

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty.$$

Consider $Y = X^2, \mathcal{Y} = (0, \infty)$. The function $g(x) = x^2$ is monotone on $(-\infty, 0)$ and on $(0, \infty)$. Take the partition

$$\begin{aligned}
&A_0 = \{0\}; \\
&A_1 = \{-\infty, 0\}, \quad g_1(x) = x^2, \quad g_1^{-1}(y) = -\sqrt{y}; \\
&A_2 = \{0, \infty\}, \quad g_2(x) = x^2, \quad g_2^{-1}(y) = \sqrt{y}.
\end{aligned}$$

The PDF of Y is

$$\begin{aligned}
f_Y(y) &= \frac{1}{\sqrt{2\pi}} e^{-(-\sqrt{y})^2/2} \left| -\frac{1}{2\sqrt{y}} \right| + \frac{1}{\sqrt{2\pi}} e^{-(\sqrt{y})^2/2} \left| \frac{1}{2\sqrt{y}} \right| \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-y/2}, \quad 0 < y < \infty.
\end{aligned}$$

This is a **Chi-Squared** random variable with 1 degree of freedom.

Average

The distribution of a random variable gives full information about the probability that the random variable will fall into any particular set.

Often time, we want one number summarizing the "average" value of the random variable.

Two ways to average a list of numbers $(1, 1, 1, 1, 1, 3, 3, 5)$:

- Arithmetic mean (*ungrouped* average)

$$\frac{1}{8}(1 + 1 + 1 + 1 + 1 + 3 + 3 + 5) = 2$$

- Weighted mean (*grouped* average)

$$\frac{5}{8} \cdot 1 + \frac{2}{8} \cdot 3 + \frac{1}{8} \cdot 5 = 2$$

Expectation of a random variable

The **expected value** (also called the **expectation** or **mean**) of a random variable X is defined by

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

provided that the integral or sum *exists*, i.e., $E(|X|) < \infty$.

We often abbreviate $E(X)$ to EX. Similarly, $E(X^2)$ to EX^2, and $E(X^n)$ to EX^n.

Bernoulli mean

A discrete random variable has a Bernoulli distribution, $X \sim \text{Bern}(p)$, if its PMF is of the form

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p. \end{cases}$$

Then

$$E(X) = 1 \cdot p + 0 \cdot (1-p) = p.$$

The frequentist interpretation:

- Consider a large number of independent Bernoulli trials, each with probability p of success
- Write 1 for "success" and 0 for "failure"
- In the long run, the proportion of 1's is very close to p, which is the average of the list of 0's and 1's

Exponential mean

Suppose X has an **Exponential distribution** with parameter λ, it has PDF given by

$$f_X(x) = \frac{1}{\lambda}e^{-x/\lambda}, \quad 0 \leq x < \infty, \quad \lambda > 0.$$

Then $E(X)$ is given by

$$\begin{aligned} E(X) &= \int_0^\infty x\frac{1}{\lambda}e^{-x/\lambda}dx \\ &= -xe^{-x/\lambda}\Big|_0^\infty + \int_0^\infty e^{-x/\lambda}dx \quad \text{[integration by parts]} \\ &= \int_0^\infty e^{-x/\lambda}dx = \lambda. \end{aligned}$$

Cauchy mean

Suppose X has a **Cauchy distribution**, its PDF is given by

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad -\infty < x < \infty.$$

This is a classic example of a random variable whose expectation does *not exist*.

Cauchy mean

Suppose X has a **Cauchy distribution**, its PDF is given by

$$f_X(x) = \frac{1}{\pi}\frac{1}{1+x^2}, \quad -\infty < x < \infty.$$

This is a classic example of a random variable whose expectation does *not exist*.

$$E|X| = \int_{-\infty}^{\infty} \frac{|x|}{\pi}\frac{1}{1+x^2}dx = \frac{2}{\pi}\int_0^{\infty} \frac{x}{1+x^2}dx.$$

For any positive number M,

$$\int_0^M \frac{x}{1+x^2}dx = \frac{\log(1+x^2)}{2}\bigg|_0^M = \frac{\log(1+M^2)}{2}.$$

Thus,

$$E|X| = \lim_{M\to\infty}\frac{2}{\pi}\int_0^M \frac{x}{1+x^2}dx = \frac{1}{\pi}\lim_{M\to\infty}\log(1+M^2) = \infty$$

and EX does not exist.

EX as a good guess at the value of X

Suppose we measure the distance between a random variable X and a constant b by $(X-b)^2$. The closer b is to X, the smaller this quantity is.

$$\begin{aligned}
E(X-b)^2 &= E(X-EX+EX-b)^2 \\
&= E((X-EX)+(EX-b))^2 \\
&= E(X-EX)^2+(EX-b)^2+2E((X-EX)(EX-b)),
\end{aligned}$$

where

$$E((X-EX)(EX-b)) = (EX-b)E(X-EX) = 0.$$

So, the expectation of the distance is

$$E(X-b)^2 = E(X-EX)^2+(EX-b)^2,$$

which is minimized by choosing $b=EX$

$$\min_b E(X-b)^2 = E(X-EX)^2.$$

Linearity of expectation

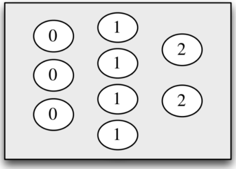
Linearity: the expected value of a sum of random variables is the sum of the individual expected values.

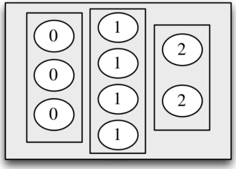
For any random variables X and Y and any constant c

$$E(X+Y) = E(X) + E(Y)$$
$$E(cX) = cE(X).$$

This holds even if X and Y are dependent.

Proof of linearity





Expectation as ungrouped average:

$$E(X) = \sum_s X(s)P(\{s\}), \quad E(Y) = \sum_s Y(s)P(\{s\})$$

Combining $\sum_s X(s)P(\{s\})$ and $\sum_s Y(s)P(\{s\})$ gives

$$\begin{aligned} E(X) + E(Y) &= \sum_s X(s)P(\{s\}) + \sum_s Y(s)P(\{s\}) \\ &= \sum_s (X+Y)(s)P(\{s\}) \\ &= E(X+Y) \end{aligned}$$

Binomial mean

For $X \sim \mathrm{Bin}(n, p)$, then

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{n} x P(X = x) \\
&= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= n \sum_{x=0}^{n} \binom{n-1}{x-1} p^x (1-p)^{n-x} \quad [x\binom{n}{x} = n\binom{n-1}{x-1}] \\
&= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^{j} (1-p)^{n-1-j} \quad [\text{sum of the } \mathrm{Bin}(n-1, p)] \\
&= np.
\end{aligned}
$$

Binomial mean with linearity

Using linearity of expectation gives a much simpler solution:

A $\mathrm{Bin}(n, p)$ random variable X can be expressed as the sum of n independent $\mathrm{Bern}(p)$ random variables

$$X = I_1 + \cdots + I_n,$$

where each I_j has expectation $E(I_j) = p$. By linearity,

$$E(X) = E(I_1) + \cdots + E(I_n) = np.$$

Expectation of $g(X)$

Linearity of expectation:

$$E(X+Y) = E(X) + E(Y) \text{ and } E(cX) = cE(X).$$

In general, $E(g(X)) \neq g(E(X))$ if g is *not linear*.

Expectation of $g(X)$

Linearity of expectation:

$$E(X+Y) = E(X) + E(Y) \text{ and } E(cX) = cE(X).$$

In general, $E(g(X)) \neq g(E(X))$ if g is *not linear*.

E.g., suppose the PMF of a random variable X is

$$f_X(0) = 1/4, \quad f_X(1) = 1/2, \quad f_X(2) = 1/4.$$

With $Z = g(X) = 2^X$,

$$f_Z(1) = 1/4, \quad f_Z(2) = 1/2, \quad f_Z(4) = 1/4.$$

Then

$$\begin{aligned} E(g(X)) &= 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} = \frac{9}{4} \\ g(E(X)) &= 2^{0 \cdot 1/4 + 1 \cdot 1/2 + 2 \cdot 1/4} = 2 \end{aligned}$$

Expectation of $g(X)$

To find $E(g(X))$:

- Find the distribution of the random variable $g(X)$
- Use the definition of expectation

Expectation of $g(X)$

To find $E(g(X))$:

- Find the distribution of the random variable $g(X)$
- Use the definition of expectation

The **law of the unconscious statistician** (LOTUS) is a powerful alternative. If X is a discrete random variable and g is a function from $\mathbb{R}$ to $\mathbb{R}$, then

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

provided that the integral or sum exists.

Proof of the LOTUS

As in the proof of linearity, the expectation of $g(X)$ can be written in ungrouped form:

$$E(g(X)) = \sum_s g(X(s))P(\{s\}),$$

where the sum is over all possible outcomes in the sample space.

We can also group the outcomes according to the value that X assigns to them:

$$\begin{aligned} E(g(X)) &= \sum_s g(X(s))P(\{s\}) \\ &= \sum_x \sum_{s:X(s)=x} g(X(s))P(\{s\}) \\ &= \sum_x g(x) \sum_{s:X(s)=x} P(\{s\}) \\ &= \sum_x g(x)P(X=x) \end{aligned}$$