# Week 8

## Statistics 251

Zhongjian Wang

Department of Statistics
The University of Chicago

# Where are we?

# Definition: Discrete Case

Recall that, for any two events $E$ and $F$, the conditional probability of $E$ given $F$ is defined, provided that $\mathbb{P}(F) > 0$, by

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(EF)}{P(F)}$$

Hence, if $X$ and $Y$ are discrete random variables, it is natural to define the conditional probability mass function of $X$ given that $Y = y$, by

$$\begin{aligned} \mathbb{P}_{X|Y}(x|y) &= \mathbb{P}\{X = x | Y = y\} \\ &= \frac{\mathbb{P}\{X = x, Y = y\}}{\mathbb{P}\{Y = y\}} \\ &= \frac{p(x, y)}{p_Y(y)} \end{aligned}$$

for all values of $y$ such that $p_Y(y) > 0$.

# Definition: Continuous Case

If $X$ and $Y$ have a joint probability density function $f(x, y)$, then the conditional probability density function of $X$ given that $Y = y$ is defined, for all values of $y$ such that $f_Y(y) > 0$, by

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

Recall if $X$, $Y$ are independent, then $f(x, y) =$
Now what is $\mathbb{P}_{X|Y}(x|y)$?

# Example: Computation

Suppose that the joint density of $X$ and $Y$ is given by

$$f(x, y) = \begin{cases} \frac{e^{-x/y}e^{-y}}{y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Find $\mathbb{P}\{X > 1 | Y = y\}$.

# Example: Bivariate Normal Distribution

One of the most important joint distributions is the bivariate normal distribution. We say that the random variables $X$, $Y$ have a bivariate normal distribution if, for constants $\mu_x$, $\mu_y$, $\sigma_x > 0$, $\sigma_y > 0$, $-1 < \rho < 1$, their joint density function is given by,

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\Big( -\frac{1}{2(1-\rho^2)} \cdot ((\frac{x-\mu_x}{\sigma_x})^2 + (\frac{y-\mu_y}{\sigma_y})^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y})\Big)$$

Proposition:

1. Given $Y = y$, the conditional distribution of $X$ isa normal distribution with mean $\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y)$ and variance $\sigma_x^2(1-\rho^2)$
2. The marginal distribution of $X$ is normal with mean $\mu_x$ and variance $\sigma_x^2$.

# Where are we?

# Conditional Expectation for Discrete r.v.

Recall that if $X$ and $Y$ are jointly discrete random variables, then the conditional probability mass function of $X$, given that $Y = y$, is defined, for all $y$ such that $P\{Y = y\} > 0$, by

$$p_{X|Y}(x \mid y) = P\{X = x \mid Y = y\} = \frac{p(x, y)}{p_Y(y)}$$

Base on this, we know $p_{X|Y}\{\cdot \mid Y = y\}$ is a distribution function.
Now how to define $E[X \mid Y = y]$?
The conditional expectation of $X$ given that $Y = y$, for all values of $y$ such that $p_Y(y) > 0$, by

$$\begin{aligned} E[X \mid Y = y] &= \sum_x x P\{X = x \mid Y = y\} \\ &= \sum_x x p_{X|Y}(x \mid y) \end{aligned}$$

# Conditional Expectation for Continuous r.v.

Similarly, let us recall that if $X$ and $Y$ are jointly continuous with a joint probability density function $f(x, y)$, then the conditional probability density of $X$, given that $Y = y$, is defined, for all values of $y$ such that $f_Y(y) > 0$, by

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$

Again, we know $p_{X|Y}\{\cdot \mid Y = y\}$ is a distribution function. Now how to define $E[X \mid Y = y]$?
The conditional expectation of $X$, given that $Y = y$, by

$$E[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) dx$$

provided that $f_Y(y) > 0$

# Example

Suppose that the joint density of $X$ and $Y$ is given by

$$f(x, y) = \frac{e^{-x/y} e^{-y}}{y} \quad 0 < x < \infty, 0 < y < \infty$$

Compute $E[X \mid Y = y]$.

# Formula

$$E[X] = E[E[X \mid Y]]$$

What does it mean?
If $Y$ is a discrete random variable,

$$E[X] = \sum_{y} E[X \mid Y = y] P\{Y = y\}$$

whereas if $Y$ is continuous random variable,

$$E[X] = \int_{-\infty}^{\infty} E[X \mid Y = y] f_Y(y) dy$$

# Example: Miner Travel

A miner is trapped in a mine containing 3 doors.

- The first door leads to a tunnel that will take him to safety after 3 hours of travel.
- The second door leads to a tunnel that will return him to the mine after 5 hours of travel.
- The third door leads to a tunnel that will return him to the mine after 7 hours.

If we assume that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until he reaches safety?

# Conditional Variance

$$\mathrm{Var}(X) = E[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(E[X \mid Y])$$

Proof can be found on Ross Book and is left as homework.

# Conditional Expectation is the best preditor-Derivation (Reading)

- Assume a random variable $X$ is observed.
- We need to predict the value of a second random variable $Y$.
- Let $g(X)$ denote the predictor for $Y$.
- We would like to show $g(X) = E[Y \mid X]$, is the best possible predictor.

How to determine a predictor is good (or best)?
One possible criterion for closeness is to choose $g$ so as to minimize $E\left[(Y - g(X))^2\right]$.

# Proof

In detail, we would like to prove,

$$E\left[(Y-g(X))^2\right] \geq E\left[(Y-E[Y \mid X])^2\right].$$

$$\begin{aligned}
E\left[(Y-g(X))^2 \mid X\right] =& E\left[(Y - E[Y \mid X] + E[Y \mid X] - g(X))^2 \mid X\right] \\
=& E\left[(Y-E[Y \mid X])^2 \mid X\right] \\
&+ E\left[(E[Y \mid X]-g(X))^2 \mid X\right] \\
&+ 2E[(Y-E[Y \mid X])(E[Y \mid X]-g(X)) \mid X]
\end{aligned}$$

However, given $X, E[Y \mid X]-g(X)$, being a function of $X$, can be treated as a constant. Thus,

$$\begin{aligned}
&E[(Y-E[Y \mid X])(E[Y \mid X]-g(X)) \mid X] \\
&\quad = (E[Y \mid X]-g(X))E[Y-E[Y \mid X] \mid X] \\
&\quad = (E[Y \mid X]-g(X))(E[Y \mid X]-E[Y \mid X]) \\
&\quad = 0
\end{aligned}$$

# Where are we?

# Markov inequalities

If $X$ is a random variable that takes only non-negative values, then, for any value $a > 0$

$$\mathbb{P}\{X \geq a\} \leq \frac{\mathbb{E}[X]}{a}$$

Proof. Consider

$$I = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases}$$

# Chebyshev's inequality

If $X$ is a random variable with finite mean $\mu$ and variance $\sigma^2$, then, for any value $k > 0$

$$\mathbb{P}\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

Proof. Note that $(X - \mu)^2$ is then a nonnegative random variable, we can apply Markov's inequality to it.

# Corollary

If $\mathrm{Var}(X) = 0$, then

$$\mathbb{P}\{X = \mathbb{E}[X]\} = 1$$

# Example: the inequalities are inaccurate

- Consider $X$ is uniformly distributed over the interval $(0, 10)$.
- What is its mean and variance? $\mathbb{E}[X] = 5$, $\mathrm{Var}(X) = \frac{25}{3}$.
- How to estimate $P\{|X - 5| > 4\}$?

# Where are we?

# Theorem

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean $E[X_i] = \mu$. Then, for any $\varepsilon > 0$,

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \to 0 \quad \text{as} \quad n \to \infty$$

# How to prove it?

- Theorem applies without any additional assumption.
- But here we assume that the random variables have a finite variance $\sigma^2$.
- Now,

$$E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu \text{ and } \text{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}$$

- Lastly, from Chebyshev's inequality that

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right\} \leq \frac{\sigma^2}{n\varepsilon^2}.$$

# Example: Concentration of Gamma r.v.

If $\{X_i\}$ are independent gamma random variables with parameters $(1,1)$, approximately how large need $n$ be so that

$$P\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - 1\right| > .01\right\} < .01?$$

# Where are we?

# Definition

The moment generating function $M(t)$ of the random variable $X$ is defined for all real values of $t$ by

$$M(t) = \mathbb{E}\left[e^{tX}\right]$$
$$= \begin{cases} \sum_x e^{tx} p(x) & \text{if } X \text{ is discrete with mass function } p(x) \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & \text{if } X \text{ is continuous with density } f(x) \end{cases}$$

# Derivaties

First,

$$\begin{aligned} M'(t) &= \frac{d}{dt}\mathbb{E}\left[e^{tX}\right] \\ &= \mathbb{E}\left[\frac{d}{dt}\left(e^{tX}\right)\right] \\ &= \mathbb{E}\left[Xe^{tX}\right] \end{aligned}$$

So,

$$M'(0) = \mathbb{E}[X].$$

What's more?

$$M^n(t) = \mathbb{E}\left[X^n e^{tX}\right] \quad n \geq 1$$

implying that

$$M^n(0) = \mathbb{E}\left[X^n\right] \quad n \geq 1$$

# Where are we?

Conditional distribution

Conditional Expectation

Inequalities for tail distribution

The weak law of large numbers

Moment Generating Function: Definitions

Moment Generating Function: Properties

# Moment Generating Function of sum of independent random variables

If $X$ and $Y$ are independent and have moment generating functions $M_X(t)$ and $M_Y(t)$, respectively. Then $M_{X+Y}(t)$, is given by

$$\begin{aligned} M_{X+Y}(t) &= E\left[e^{t(X+Y)}\right] \\ &= E\left[e^{tX}e^{tY}\right] \\ &= E\left[e^{tX}\right]E\left[e^{tY}\right] \\ &= M_X(t)M_Y(t). \end{aligned}$$

If $\{X_i\}_{i=1}^n$ are identical independent random variables from the same distribution with moment generating function $M_X$, what is $M_{\sum_{i=1}^n X_i}$?

# Lemma: moment generating function decides the distribution

- Let $Z_1, Z_2, \dots$ be a sequence of random variables having distribution functions $F_{Z_n}$ and moment generating functions $M_{Z_n}, n \geq 1$.
- Let $Z$ be a random variable having distribution function $F_Z$ and moment generating function $M_Z$.
- If $M_{Z_n}(t) \to M_Z(t)$ for all $t$, then $F_{Z_n}(t) \to F_Z(t)$ for all $t$ at which $F_Z(t)$ is continuous.