Sarah Cooper

Presentation 2

Week 4 Question 2

2. **Expectation for Transformations** (Blitzstein 4.5, Rice 4.1-4.2, 4.6,)

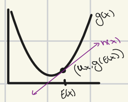
   a) Point out that, for a non-linear function $g(x)$, it is unusual for $E(g(X))$ to equate to $g(E(X))$. Give justification for Jensen's inequality and state its implications for expected values. For example, show that $E(1/X) \geq 1/E(X)$, and that $E(\log(X)) \leq \log(E(X)) \leq E(X) - 1$.

Jensen's Inequality:

a) - when g is convex:

$$g(E(x)) \leq E(g(x))$$

Convex



tangent line: $h(x) = g(E(x)) + g'(E(x))(x - E(x))$

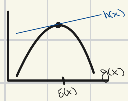
$g(x) \geq h(x) \quad \therefore \quad E(g(x)) \geq E(h(x)) = g(E(x))$

So, $E(g(x)) \geq g(E(x))$

- when g is concave

$$g(E(x)) \geq E(g(x))$$

Concave



Here $E(g(x)) \leq E(h(x)) = g(E(x))$

So $E(g(x)) \leq g(E(x))$

show $E(\frac{1}{x}) \geq \frac{1}{E(x)}$.

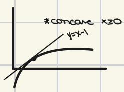
$g(x) = \frac{1}{x}$

$g'(x) = -\frac{1}{x^2} \quad g''(x) = \frac{2}{x^3} \therefore$ convex

then use Jensen's shows that

$E(\frac{1}{x}) \geq \frac{1}{E(x)}$

show that $E(\log(x)) \leq \log(E(x)) \leq E(x) - 1$



$g(x) = \log(x)$

$g'(x) = \frac{1}{x} \quad g''(x) = \frac{-1}{x^2} \therefore$ concave so $E(\log(x)) \leq \log(E(x))$

second part: $\log(x) \leq x - 1$

$\log(E(x)) \leq E(x) - 1$

so $E(\log(x)) \leq \log(E(x)) \leq E(x) - 1$

b) State Theorem A of 4.1 in Rice and explain why this is called the law of the unconscious statistician (see also Blitzstein 4.5). Explain how LOTUS is used when we find $\mathrm{Var}(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$. Describe what would be involved to find a variance without using LOTUS (e.g., for a Gamma variable).

b) Theorem A:

Suppose $Y = g(x)$

a) If x is discrete with frequency function $p(x)$, then

$$E(Y) = \sum_x g(x)p(x) \quad \text{provided that} \quad \sum |g(x)|p(x) < \infty$$

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

$$E(Y) = \int_{-\infty}^{\infty} g(x)f(x)\,dx \quad \text{provided that} \quad \int |g(x)|f(x)\,dx < \infty$$

Why the name? You can get the $E(g(x))$ knowing only frequency function of X ($p(x)$).
We do not need the PMF of $g(x)$.
When going from $E(x)$ to $E(g(x))$ it is tempting to change x to $g(x)$ in the definition, which can be done easily in an unconscious state. Sounds too good to be true, but LOTUS says it is true.

LOTUS for Variance:

$\mathrm{Var}(x) = E(x^2) - (E(x))^2$

$Y = x^2 \rightarrow$ find $f_Y(y)$
$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)\,dy$

with out LOTUS you need to find a new pdf $f_Y(y)$

but w/ LOTUS $E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx$

Finding variance for Gamma

$X \sim \mathrm{Gamma}(k, \lambda)$

$f_X(x) = \dfrac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)}$ $\quad x \geq 0 \quad$ we know $E(X) = \frac{k}{\lambda}$ and $\mathrm{Var}(x) = \frac{k}{\lambda^2}$

To find Variance w/out LOTUS you use moment generating functions

First Moment: $E(X) = \frac{k}{\lambda}$

Second Moment: $E(x^2) = \int_0^{\infty} x^2 f_X(x)\,dx = \int_0^{\infty} \dfrac{\lambda^k e^{-\lambda x} x^{k+1}}{\Gamma(k)}\,dx = \dfrac{\lambda^k}{\Gamma(k)} \int_0^{\infty} e^{-\lambda x} x^{k+1}\,dx = \dfrac{\lambda^k}{\Gamma(k)}\left(\dfrac{\Gamma(k+2)}{\lambda^{k+2}}\right) = \dfrac{k(k+1)}{\lambda^2}$

$\mathrm{Var}(x) = E(x^2) - E(x)^2 = \dfrac{k(k+1)}{\lambda^2} - \left(\dfrac{k}{\lambda}\right)^2 = \dfrac{k}{\lambda^2}$

c) State the multivariate version of LOTUS and show this implies $E(X+Y) = E(X)+E(Y)$, even if X and Y are not independent. Describe what would be involved to find a covariance without using LOTUS

c) Multivariate LOTUS: Rice 4.1 Theorem B

Suppose that $X_1, \ldots, X_n$ are jointly distributed random variables and $Y = g(X_1, \ldots, X_n)$.

a. If the X_i are discrete with frequency function $p(x_1, \ldots, x_n)$, then

$$E(Y) = \sum_{x_1,\ldots,x_n} g(x_1, \ldots, x_n)p(x_1, \ldots, x_n)$$

provided that $\sum_{x_1,\ldots,x_n} |g(x_1, \ldots, x_n)|p(x_1, \ldots, x_n) < \infty$.

b. If the X_i are continuous with joint density function $f(x_1, \ldots, x_n)$, then

$$E(Y) = \int\int\cdots\int g(x_1, \ldots, x_n)f(x_1, \ldots, x_n)dx_1\cdots dx_n$$

provided that the integral with $|g|$ in place of g converges.

show that $E(X+Y) = E(X) + E(Y)$ despite independence

$g(x,y) = X+Y$ → transformation function

$E(X+Y) = \int\int_{-\infty}^{\infty} (x+y)\, f_{X,Y}(x,y)\, dx\, dy$

$= \int_{-\infty}^{\infty} x \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy}_{\text{marginal distribution of } X} dx + \int_{-\infty}^{\infty} y \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx}_{\text{marginal distribution of } Y} dy$

$E(X+Y) = \int_{-\infty}^{\infty} x f_X(x)dx + \int_{-\infty}^{\infty} y f_Y(y)dy$

$E(X+Y) = E(X) + E(Y)$

Covariance

$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(XY) - E(X)E(Y)$

$\left.\begin{array}{l} E(XY) = \int\int xy\, f_{X,Y}(x,y)dxdy \end{array}\right\}$ LOTUS

$E(X) = \int x f_X(x)dx$

$E(Y) = \int y f_Y(y)dy$

$E(XY) - E(X)E(Y)$

without LOTUS

Define $Z = XY$ find $f_Z(z)$

then $E(XY) = E(Z) = \int_{-\infty}^{\infty} z f_Z(z)dz$

then also find $E(X)$ and $E(Y)$ through two more pdfs.

d) Show that, for uncorrelated random variables, $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$, but that this is not true in general. Give examples using extremely correlated variables (e.g., $X_1 = X_2$ and $X_1 = -X_2$).

d) Show that Uncorrelated : $Var(x+Y) = Var(x) + Var(Y)$

$$Var(x+Y) = E[(x+Y - E(x+Y))^2]$$

$$Var(x+Y) = E[(x - E(x) + Y - E(Y))^2]$$

$$= \underbrace{E(x-\mu_x)^2}_{Var(x)} + \underbrace{2(x-E(x))(Y-E(Y))}_{2Cov(x,Y)} + \underbrace{(Y-E(Y))^2}_{Var(Y)}$$

$$= Var(x) + 2Cov(x,Y) + Var(Y)$$

when uncorrelated $Cov(x,Y) = 0$

so $Var(x+Y) = Var(x) + Var(Y) + 2(0)$

$= Var(x) + Var(Y)$

Does not hold:

Ex) $X_1 = X_2$

Then $Var(X_1+X_2) = Var(X_1+X_1) = Var(2X_1) = 4Var(X_1)$

↳ that is only equal to $Var(X_1) + Var(X_2)$ when $Var(X_1) = 0$

Ex2) $X_1 = -X_2$

$Var(X_1+X_2) = Var(X_1 - X_1) = Var(0) = 0$

But $Var(X_1) + Var(X_2) = Var(X_1) + Var(-X_1) = Var(X_1) + Var(X_1) = 2Var(X_1)$

↳ once again only holds if $Var(X_1) = 0$

e) Show how to find approximations to the mean and variance of a transformation of a random variable using Taylor's approximation. As an example, find approximations to the mean and variance of $Y = \log(X)$, for $X \sim \text{Gamma}(\alpha, \lambda)$.

e) $Y = g(x) \approx \underbrace{g(\mu_x)}_{\text{constant}} + \underbrace{(x-\mu_x)g'(\mu_x)}_{0} + \frac{(x-\mu_x)^2}{2} g''(\mu_x)$

$E(g(x)) = g(\mu_x) + 0 + \frac{Var(x)}{2} g''(\mu_x)$

Example) $x \sim Gamma(\alpha, \lambda)$

$Y = \log(x)$ $\quad E(X) = \frac{\alpha}{\lambda}$ $\quad Var(x) = \frac{\alpha}{\lambda^2}$ $\quad g(x) = \log(x)$

$g'(x) = \frac{1}{x}$ $\quad g''(x) = \frac{-1}{x^2}$

$E(\log(x)) = \log\left(\frac{\alpha}{\lambda}\right) + \frac{\alpha/\lambda^2}{2}\left(\frac{-1}{(\alpha/\lambda)^2}\right)$

$= \log\left(\frac{\alpha}{\lambda}\right) - \frac{1}{2\alpha}$

Recall $E(\log(x)) \leq \log(E(x))$ (from Jensen's)

so since we are subtracting from $\log(\frac{\alpha}{\lambda})$ this inequality holds true.

$Var(g(x)) \approx Var(g(\mu_x) + (x-\mu_x)g'(\mu_x))$

$= Var(x\, g'(\mu_x))$

$= (g'(\mu_x))^2\, Var(x)$ } general explanation

$Var(\log(x)) = \left(\frac{1}{\alpha/\lambda}\right)^2\left(\frac{\alpha}{\lambda^2}\right) = \frac{1}{\alpha}$