Stat 111

Olivia R. McClammy

February 12, 2025

For this problem, I made references to "Introduction to Probability" by Joseph K. Blitzstein and Jessica Hwang and "Mathematical Statistics: Third Edition" by John A. Rice. I would also like to thank Professor Everson on his instruction and for discussion with Sarah Cooper on this problem.

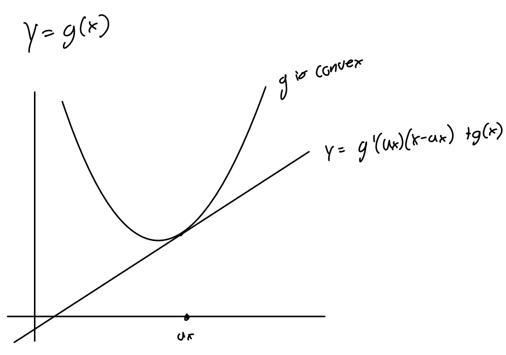
1 Question 2

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 $E(log(X)) \leq log(E(X)) \leq E(X) - 1$.

For a non-linear function $g(X)$, it is unusual for $E(g(X)) = g(E(X))$

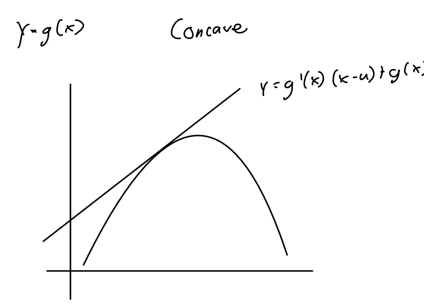
Convex = second derivative is positive
Concave = second derivative is negative

Consequences of convex and concave:
For a convex function, the tangent line is equal or below the convex function.
For a concave function, the tangent line is equal or above the concave function.



$g(x) \geq g(u_x)$

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

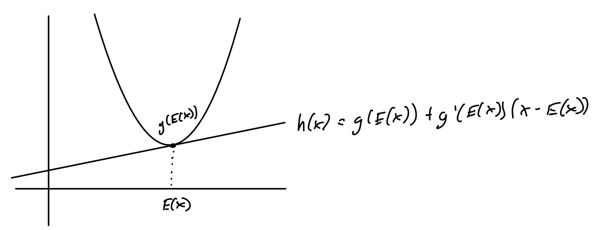


$g(x) \leq g(u_x)$

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

Jensen's Inequality (From Blitzstein - page 425): Let X be a random variable. If g is a convex function, then $E(g(X)) \geq g(E(X))$. If g is a concave function, then $E(g(X)) \leq g(E(X))$. In both cases, the only way that the equality holds is if there are constants a and b such that $g(X) = a + bX$ with probability 1.

Let $g(x) = x^2$



Then $g(x) \geq h(x)$

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

$E(h(x)) = g(E(x)) + 0$

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

$E(\underline{1/X}) \geq 1/E(X)$

$g(x) = 1/x$

$g'(x) = -1/x^2$

$g''(x) = 2/x^3 \leftarrow g$ is convex

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

$E(1/x) \geq 1/E(x)$

$\underbrace{E(\underline{log}(X)) \leq log(E(X))}\underbrace{\leq E(X) - 1}$

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

$g'(x) = 1/x$

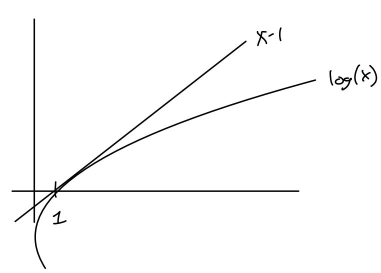
$g''(x) = -1/x^2$

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

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

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

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



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

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 $Var(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$. Describe what would be involved to find a variance without using LOTUS (e.g., for a Gamma variable).

Theroem A: Rice (page 122)

Suppose that $Y = g(X)$

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

$E(Y) = \sum_x g(X)p(X)$

provided that $\sum |g(X)|p(X)$

b.) If X is continous with density function $f(X)$, then

$E(Y) = \int_{-\infty}^{\infty}(g(X)f(X)dx$

provided that $\int |g(X)|f(X)dx$

We call this the law of unconscious statistician;'s law because we do not think about using it and we use LOTUS all the time.

How is LOTUS used in $Var(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$?

When we use LOTUS:

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f_x(x)dx$$

If we don't use LOTUS

$$Y = x^2$$

$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)dy$, where we would have to find the pdf of Y.

$$\begin{aligned} Var(x) &= E((x-E(x))^2) = E(x^2) - E(x)^2 \\ &= E(x^2 - 2(E(x)(x) + E(x)^2) \\ &= E(x^2) - E(2(E(x)(x)) + E(E(x)^2) \\ &= E(x^2) - 2E(x)^2 + E(x)^2 \\ &= E(x^2) - E(x)^2 \end{aligned}$$

Suppose that $Y = g(X)$. If we did not have LOTUS, we would need to find the pdf for Y and evaluate the integral of y multiplied by y's pdf.

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.

Multivariate Version of LOTUS: Rice (page 123)

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

a.) If the X_i are discrete with frequency function $p(X_1, ..., X_n)$, then

$$E(Y) = \sum_{X_1,...,X_n} g(X_1, ..., X_n)p(X_1, ..., X_n)$$

provided that $\sum_{X_1,...,X_n} |g(X_1, ..., X_n)|p(X_1, ..., X_n) < \infty$

b.) If the X_i are continuous with joint density function $f(X_1, ..., X_n)$, then

$$E(Y) = \int\int ... \int g(X_1, ..., X_n)f(X_1, ..., X_n)dX_1, ...dX_n$$

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

How does the multivariate version of LOTUS imply $E(X+Y) = E(X) + E(Y)$?

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

Let $T = x+y$

Find the pdf T, $E(x+y) = \int t\, f_T(t)\,dt$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x f_{xy}(x,y) + y\, f_{xy}(x,y)\right)dx\,dy$$

$$\underbrace{\int_{-\infty}^{\infty} x \int_{-\infty}^{\infty} f_{xy}(x,y)\,dy\,dx}_{E(x)} + \underbrace{\int_{-\infty}^{\infty} y \int_{-\infty}^{\infty} f_{xy}(x,y)\,dx\,dy}_{E(y)}$$

$$E(x+y) = E(x) + E(y)$$

If we did not have LOTUS, to find the covariance between X and Y, we would have to use density $Z = X+Y$ and find the pdf of Z, and then find $\int_{-\infty}^{\infty} f(Z)ZdZ = E(X+Y)$, we would also have to find the density of $E(T)$ where $T = X \cdot Y$

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

$Var(X+Y) = Var(X) + Var(Y)$

Let $Z = X+Y$

$$Var(Z) = E(Z^2) - E(Z)^2$$

$$= E((X+Y)^2) - (E(X+Y))^2$$

$$= E(X^2 + 2XY + Y^2) - \underbrace{E(X+Y)^2}$$

$$= \underbrace{E(X^2) - E(X)^2}_{Var(X)} + \underbrace{E(Y^2) - E(Y)^2}_{Var(Y)} + \underbrace{2(E(XY) - E(X)E(Y))}_{Covariance}$$

$$Var(X+Y) = Var(X) + Var(Y) + 2(E(XY) - E(X)E(Y))$$ If independent = $2Var(X_1)$

If $X_1 = X_2, \rho = 1$

$$Var(X_1 + \underbrace{X_2}_{X_1}) = Var(2X_1) = 2^2 Var(X_1) = 4Var(X_1)$$

If $X_1 = -X_2, \rho = -1$

$$Var(\underbrace{X_1}_{-X_2} + X_2)$$

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

$$Var(0) = 0$$

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 $X - Gamma(\alpha, \lambda)$.

Taylor's approximation: $g(x) = g(\mu_x) + \underbrace{(x-\mu_x)g'(x)}_{\mu_x - \mu_x} + ((x-\mu_x)/2)g''(x) + \ldots$

$Y = log(x)$, for $X\tilde{}Gamma(\alpha, \lambda)$

Find an approximation to the mean:

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

$$g'(x) = 1/x \quad , \quad g''(x) = -1/x^2$$

$$\text{Gamma} = u_x = \frac{\alpha}{\lambda} \; , \quad Var(x) = \frac{\alpha}{\lambda^2}$$

$$= g(u_x) + \underbrace{0} + \frac{Var(x)}{2} g''(u_x)$$

$$= \log(g(u_x)) + \log\left(\frac{Var(x)}{2}\right) g''(u_x)$$

$$= \log(\alpha/\lambda) + -1/2 \left(\frac{\alpha}{\lambda^2}\right)\left(\frac{\alpha^2}{\lambda^2}\right) = \underbrace{\log(\alpha/\lambda) - 1/2\alpha \leq \log(E(x))}_{\text{Jensen's inequality}}$$

Find an approximation for variance:

We want to use first order approximation

$$g(x) \approx \underbrace{g(u_x)} + \underbrace{(x-u_x)}\, \underbrace{g'(u_x)}$$

$$Var(g(x)) \approx Var(x)\,(g'(u_x))^2$$

$$= \left(\alpha/\lambda^2\right)\left(\lambda^2/\alpha^2\right) = 1/\alpha$$