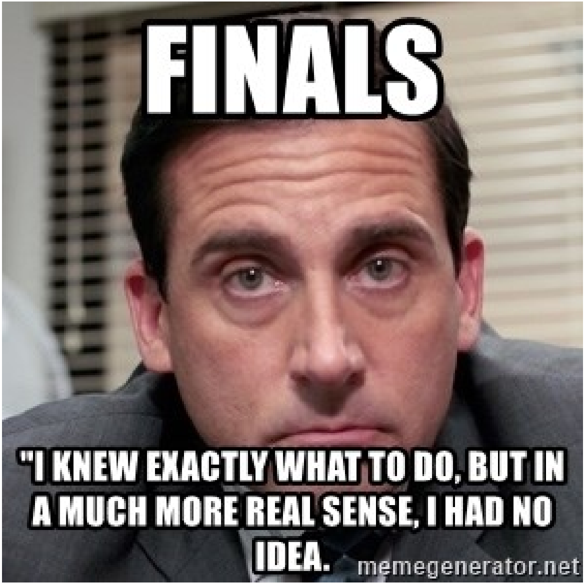
FINALS
"I KNEW EXACTLY WHAT TO DO, BUT IN A MUCH MORE REAL SENSE, I HAD NO IDEA.
memegenerator.net

Today's Agenda

Last time:

- Central Limit Theorem
- Moment generating functions

. . . we finished the course material!

Today (Lec 36, 07/26):

- Practice
- Course evaluation

Question 1

$$X = \begin{cases} 1 & \text{with prob } p \\ 0 & \text{— " —} \quad 1-p \end{cases}$$

Suppose $X \sim \mathit{Bernoulli}(p)$ for some $p \in (0,1)$. Compute the mgf of X and use it to compute $E(X)$ and $Var(X)$.

$$M(t) = E(e^{tX}) = e^{t\cdot 0} P(X=0) + e^{t\cdot 1} P(X=1)$$

$$= 1-p + pe^{t} \quad \text{for } t \in \mathbb{R}$$

$$E(X) = M'(0) = (p\cdot e^{t})\Big|_{t=0} = p$$

$$E(X^2) = M''(0) = (p\cdot e^{t})\Big|_{t=0} = p$$

$$\Rightarrow Var(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p)$$

Question 2

Let X, Y be arbitrary random variables, which of the following is always correct?

- If X and Y are independent, they are uncorrelated. ✓
- If X and Y are uncorrelated, they are independent. ✗
- If X and Y are dependent, they are correlated. ✗

if $\text{corr} \neq 0 \Rightarrow X, Y$ dependent

Question 3

In a laboratory, the temperature of an object is measured independently with two thermometers A and B. Let X and Y show the temperatures of the object measured by thermometers A and B, respectively. We decide to combine the two measurements as a weighted average of X and Y to report the temperature of the object, i.e. the reported temperature T is $T = \alpha X + (1 - \alpha) Y$ where $\alpha \in (0, 1)$. For what value of α is the variance of the random variable T minimized?

We compute $\mathrm{Var}(T)$ and minimize it wrt α.

$$\mathrm{Var}(T) = \mathrm{Var}(\alpha X + (1-\alpha)Y)$$

$$= \alpha^2 \mathrm{Var}(X) + (1-\alpha)^2 \mathrm{Var}(Y) + 2\alpha(1-\alpha) \underbrace{\mathrm{Cov}(X,Y)}_{=0 \text{ bc } X,Y \text{ indep}}$$

$$= \alpha^2 \cdot (\mathrm{Var}(X) + \mathrm{Var}(Y)) - 2\alpha \mathrm{Var}(Y) + \mathrm{Var}(Y)$$

We set the 1st derivative zero and solve for α:

$$\frac{d}{d\alpha} \mathrm{Var}(T) = 2\alpha (\mathrm{Var}(X) + \mathrm{Var}(Y)) - 2\mathrm{Var}(Y) \overset{!}{=} 0$$

$$\Rightarrow \quad \alpha^* = \frac{\mathrm{Var}(Y)}{\mathrm{Var}(X) + \mathrm{Var}(Y)}$$

We check the 2nd derivative to make sure it's a minimum:

$$\frac{\partial^2}{\partial \alpha^2} \mathrm{Var}(T) = 2(\mathrm{Var}(X) + \mathrm{Var}(Y)) > 0$$

$$\Longrightarrow \quad \alpha^* = \frac{\mathrm{Var}(Y)}{\mathrm{Var}(X) + \mathrm{Var}(Y)} \text{ minimizes } \mathrm{Var}(T).$$

Question 4

Let X be a continuous random variable with pdf

$$f(x) = e^{-x}, \quad x \geq 0$$

50% quantile

and 0 otherwise. Compute the median of X.

The cdf of X is

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x} e^{-t}\,dt = 1 - e^{-x}$$

Solve $F(x) = 0.5$ to find median.

that's the median

$$F(x) = 1/2 \iff e^{-x} = \frac{1}{2} \iff \underline{x = \log(2)}$$

Question 5

It is known that 8% of the world's population has blue eyes. Suppose you would like to estimate this proportion by asking n people wether they have blue eyes. You would use as an estimate X/n where X is the number of people having blue eyes in a sample of n people. How many people do you need to ask so that with probability 95%, the estimate will be within 0.01 of its true value 0.08?

We need to find $n \in \mathbb{N}$ such that

$$P\left(\left|\frac{X}{n} - 0.08\right| \leq 0.01\right) \geq 0.95$$

where $X \sim \text{Bin}(n, 0.08)$

$$0.95 \overset{!}{=} P\left(|X - 0.08n| \le 0.01\, n \right)$$

$$= P\left(\frac{|X - 0.08n|}{\sqrt{0.08 \cdot 0.92 n}} \le \frac{0.01\, n}{\sqrt{0.08 \cdot 0.92 n}} \right)$$

$$\overset{CLT}{\approx} P\left(|Z| \le 0.1359\sqrt{n} \right).$$

We know $P(|Z| \le 1.96) = P(-1.96 \le Z \le 1.96)$
$\qquad\qquad\qquad\qquad\qquad\qquad = 0.95$

SO $\quad 0.1359\sqrt{n} \ge 1.96 \implies n \ge \underline{60.83}.$

$\implies$ <u>n at least 61.</u>

Question 6

Suppose that $X_1, ..., X_{50}$ are independent Geometric random variables with parameter 0.5. Estimate the probability that $\sum_{i=1}^{50} X_i > 55$.

Note $E(X_i) = \frac{1/2}{1/2} = 1$, $Var(X_i) = \frac{1/2}{(1/2)^2} = 2$

and the X_i are indep.

Thus, $\sum_{i=1}^{50} X_i \overset{approx}{\sim} N(50, 100)$ by CLT.

So that

$$P\left(\sum_{i=1}^{50} X_i > 55\right) \overset{\text{cont corr}}{=} P\left(\sum_{i=1}^{50} X_i \geq 55.5\right)$$

$$\overset{CLT}{\approx} P\left(Z \geq \frac{55.5-50}{\sqrt{100}}\right) = 1 - P(Z \leq 0.55)$$

$$= \underline{\underline{0.29116}}$$

Question 7

Let X be a continuous rv with pdf

$$f(x) = \lambda x^{\lambda - 1}, \quad 0 < x < 1,$$

and 0 otherwise.

a) Compute $E(X^k)$ for $k \in \mathbb{N}$.

b) Let $Y = -\lambda \log(X)$. Determine the pdf of Y.

a) $E(X^k) = \int_{-\infty}^{\infty} x^k f(x)\,dx = \int_0^1 x^k \lambda x^{\lambda-1}\,dx$ (note the support)

$$= \lambda \int_0^1 x^{k+\lambda-1}\,dx = \lambda \left[\frac{x^{k+\lambda}}{k+\lambda}\right]_0^1 = \underline{\underline{\frac{\lambda}{k+\lambda}}}$$

for $k \in \mathbb{N}$.

b) Let $Y = -\lambda \log(X)$. We first compute the cdf of X:

$$F_X(x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^x \lambda t^{\lambda-1}\,dt = x^\lambda \quad \text{for } x \in [0,1]$$

Using this we find the cdf of Y as

$$F_Y(y) = P(Y \le y) = P(-\lambda \log(X) \le y) = P\left(\log(X) \ge -\frac{y}{\lambda}\right)$$

$$= P(X \ge e^{-y/\lambda}) = 1 - F_X(e^{-y/\lambda}) = 1 - e^{-y}, \quad y > 0$$

We find the pdf of Y by taking the derivative:

$$f_Y(y) = F_Y'(y) = e^{-y}, \quad y > 0.$$

Course Perceptions

Please go to `https://perceptions.uwaterloo.ca` and fill out the survey.