Lecture 7: Joint Distributions and the Law of Large Numbers

Sta230/Mth230

Colin Rundel

February 7, 2014

Practice Problem - Skewness of Bernoulli Random Variable

Let $X \sim \text{Bern}(p)$ We have shown that

$$E(X) = p$$
$$Var(X) = p(1-p)$$

Find the Skewness of X where skewness is defined as

$$E\left(\left(\frac{X-E(X)}{SD(X)}\right)^3\right) = \frac{E\left((X-\mu)^3\right)}{\sigma^3}$$

Joint Distributions - Example

Draw two socks at random, without replacement, from a drawer full of twelve colored socks:

6 black, 4 white, 2 purple

Let B be the number of Black socks, W the number of White socks drawn, then the distributions of B and W are given by:

	0	1	2
P(B=k)	$\frac{6}{12}\frac{5}{11}=\frac{15}{66}$	$2\frac{6}{12}\frac{6}{11}=\frac{36}{66}$	$\frac{6}{12}\frac{5}{11}=\frac{15}{66}$
P(W=k)	$\frac{8}{12}\frac{7}{11}=\frac{28}{66}$	$2\frac{4}{12}\frac{8}{11}=\frac{32}{66}$	$\frac{4}{12}\frac{3}{11}=\frac{6}{66}$

Note - $B \sim \text{HyperGeo}(12,6,2) = \frac{\binom{6}{k}\binom{6}{2-k}}{\binom{12}{2}}$ and $W \sim \text{HyperGeo}(12,4,2) = \frac{\binom{4}{k}\binom{8}{2-k}}{\binom{12}{2}}$

Joint Distributions - Example, cont.

Let B be the number of Black socks, W the number of White socks drawn, then the distributions of B and W are given by:

		W = 0	W = 1	W = 2	
B	0	$\frac{1}{66}$	$\frac{8}{66}$	$\frac{6}{66}$	$\frac{15}{66}$
	1	$\frac{12}{66}$	$\frac{24}{66}$	0	$\frac{36}{66}$
	2	$\frac{15}{66}$	0	0	$\frac{15}{66}$
		$\frac{28}{66}$	$\frac{32}{66}$	$\frac{6}{66}$	$\frac{66}{66}$

$$P(B=b, W=w) = \frac{\binom{6}{b}\binom{4}{w}\binom{2}{2-b-w}}{\binom{12}{2}}$$

Marginal Distribution

Note that the column and row sums are the distributions of B and W respectively.

$$P(B=b) = P(B=b, W=0) + P(B=b, W=1) + P(B=b, W=2)$$
$$P(W=w) = P(B=0, W=w) + P(B=1, W=w) + P(B=2, W=w)$$

These are the *marginal* distributions of B and W.

In general,

$$P(X=x) = \sum_{\text{all } y} P(X=x, Y=y) = \sum_{\text{all } y} P(X=x|Y=y)P(Y=y)$$
$$= \int_{\text{all } y} P(X=x, Y=y)\ dy = \int_{\text{all } y} P(X=x|Y=y)P(Y=y)\ dy$$

Conditional Distribution

Conditional distributions are defined as we have seen previously with

$$P(X=x|Y=y) = \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{\text{joint}}{\text{marginal}}$$

Therefore the pmf for white socks given no black socks were drawn is

$$P(W=w|B=0) = \frac{P(W=w, B=0)}{P(B=0)} = \begin{cases} \frac{1}{66} \Big/ \frac{15}{66} = \frac{1}{15} & \text{if } W=0 \\ \frac{8}{66} \Big/ \frac{15}{66} = \frac{8}{15} & \text{if } W=1 \\ \frac{6}{66} \Big/ \frac{15}{66} = \frac{6}{15} & \text{if } W=2 \end{cases}$$

Expectation of Discrete Joint Distributions

$$E\big(g(X,Y)\big) = \sum_x \sum_y g(x,y) P(X=x, Y=y)$$

For example we can define $g(x,y) = x \cdot y$ then

$$\begin{aligned} E(BW) =& (0\cdot 0\cdot 1/66) + (0\cdot 1\cdot 8/66) + (0\cdot 2\cdot 6/66) \\ &+ (1\cdot 0\cdot 12/66) + (1\cdot 1\cdot 24/66) + (1\cdot 2\cdot 0/66) \\ &+ (2\cdot 0\cdot 15/66) + (2\cdot 1\cdot 0/66) + (1\cdot 2\cdot 0/66) \\ =& 24/66 = 4/11 \end{aligned}$$

Note that $E(BW) \neq E(B)E(W)$ since

$$\begin{aligned} E(B)E(W) =& \ (0\cdot 15/66 + 1\cdot 36/66 + 2\cdot 15/66) \\ & \times (0\cdot 28/66 + 1\cdot 32/66 + 2\cdot 6/66) \\ =& \ 66/66 \times 44/66 = 2/3 \end{aligned}$$

This implies that B and W are not independent and $Cov(X,Y) \neq 0$.

Expectation of Discrete Conditional Distribution

Works like any other discrete distribution

$$E(X|Y=y) = \sum_x x\ P(X=x|Y=y)$$

Therefore we can calculating things like conditional means and variances,

$$E(W|B=0) = 0\cdot 1/15 + 1\cdot 8/15 + 2\cdot 6/15 = 20/15 = 1.333$$

$$E(W^2|B=0) = 0^2\cdot 1/15 + 1^2\cdot 8/15 + 2^2\cdot 6/15 = 32/15 = 2.1333$$

$$\begin{aligned} Var(W|B=0) &= E(W^2|B=0) - E(W|B=0)^2 \\ &= 32/15 - (4/3)^2 = 16/45 = 0.3556 \end{aligned}$$

Joint Distribution - Example

Suppose that X and Y have a discrete joint distribution for which the joint pmf is defined as follows:

$$f(x,y) = \begin{cases} c|x+y| & \text{for } x,y \in \{-2,-1,0,1,2\} \\ 0 & \text{otherwise} \end{cases}$$

a) What is the value of the constant c

b) $P(X = 0 \text{ and } Y = -2)$

c) $P(X = 1)$

d) $P(X = -1 | Y = 0)$

e) $P(|X - Y| \leq 1)$

From De Groot and Schervish (2011)

Joint Distribution - Example

Suppose that X and Y have a discrete joint distribution for which the joint pmf is defined as follows

$$f(x,y) = \begin{cases} \frac{1}{30}(x+y) & \text{for } x = 0,1,2 \text{ and } y = 0,1,2,3 \\ 0 & \text{otherwise} \end{cases}$$

a) Determine the marginal pmf's of X and Y.

b) Are X and Y independent?

From De Groot and Schervish (2011)

Multinomial Distribution

Let $X_1, X_2, \cdots, X_k$ be the k random variables that reflect the number of outcomes belonging to category k in n trials with the probability of success for category k being p_k, $X_1, \cdots, X_k \sim \text{Multinom}(n, p_1, \cdots, p_k)$

$$\begin{aligned} P(X_1 = x_1, \cdots, X_k = x_k) &= f(x_1, \cdots, x_k | n, p_1, \cdots, p_k) \\ &= \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} \end{aligned}$$

$$\text{where } \sum_{i=1}^{k} x_i = n \text{ and } \sum_{i=1}^{k} p_i = 1$$

$$\begin{aligned} E(X_i) &= np_i \\ Var(X_i) &= np_i(1-p_i) \\ Cov(X_i, X_j) &= -np_i p_j \end{aligned}$$

Multinomial Example

Some regions of DNA have an elevated amount of GC relative to AT base pairs. If in a normal region of DNA we expect equal amounts of ACGT vs a GC rich region which has twice as much GC as AT. If we observe the following sequence ACTGACTTGGACCCGACGGA what is the probability that it is from a normal region or a GC rich region.

Markov's Inequality

For any random variable $X \geq 0$ and constant $a > 0$, then

$$P(X \geq a) \leq \frac{E(X)}{a}$$

Corollary - Chebyshev's Inequality:

$$P(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2}$$

"The inequality says that the probability that X is far away from its mean is bounded by a quantity that increases as $Var(X)$ increases."

Derivation of Markov's Inequality

Let X be a random variable such that $X \geq 0$ then

Derivation of Chebyshev's Inequality

Proposition - if $f(x)$ is a non-decreasing function then

$$P(X \geq a) = P\Big(f(X) \geq f(a)\Big) \leq \frac{E\big(f(X)\big)}{f(a)}$$

If we define the positive valued random variable to be $|X - E(X)|$ and $f(x) = x^2$ then

Chebyshev's Inequality - Example

Use Chebyshev's inequality to make a statement about the bounds for the probability of being with in 1, 2, or 3 standard deviations of the mean for all random variables.

If we define $a = k\sigma$ where $\sigma = \sqrt{Var(X)}$ then

$$P(|X - E(X)| \geq k\sigma) \leq \frac{Var(X)}{k^2\sigma^2} = \frac{1}{k^2}$$

Independent and Identically Distributed (iid)

A collection of random variables that share the same probability distribution and all are mutually independent.

Example

If $X \sim \text{Binom}(n,p)$ then $X = \sum_{i=1}^{n} Y_i$ where $Y_1, \cdots, Y_n \overset{iid}{\sim} \text{Bern}(p)$

Sums of iid Random Variables

Let $X_1, X_2, \cdots, X_n \overset{iid}{\sim} D$ where D is some probability distribution with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. We defined $S_n = X_1 + X_2 + \cdots + X_n$

$$\begin{aligned} E(S_n) &= E(X_1 + X_2 + \cdots + X_n) \\ &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= \mu + \mu + \cdots + \mu = n\mu \end{aligned}$$

$$\begin{aligned} Var(S_n) &= E[((X_1 + X_2 + \cdots + X_n) - (\mu + \mu + \cdots + \mu))^2] \\ &= E[((X_1 - mu) + (X_2 - \mu) + \cdots + (X_n - \mu))^2] \\ &= \sum_{i=1}^{n} E[(X_i - \mu)^2] + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} E[(X_i - \mu)(X_j - \mu)] \\ &= \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j}}^{n} Cov(X_i, X_j) = n\sigma^2 \end{aligned}$$

Average of iid Random Variables

Let $X_1, X_2, \cdots, X_n \overset{iid}{\sim} D$ where D is some probability distribution with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$.

We defined $\overline{X}_n = (X_1 + X_2 + \cdots + X_n)/n$ then

$$E(\bar{X}_n) = E(S_n/n) = E(S_n)/n = \mu$$

$$\begin{aligned} Var(\bar{X}_n) &= Var(S_n/n) \\ &= \frac{1}{n^2} Var(S_n) \\ &= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

Weak Law of Large Numbers

Based on these results and Markov's Inequality we can show the following:

Therefore, as long as $\sigma^2 < \infty$

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0 \quad \Rightarrow \quad \lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1$$

Law of Large Numbers

Weak Law of Large Numbers ($\bar{X}_n$ converges in probability to μ):

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \epsilon) = 0$$

Strong Law of Large Numbers ($\bar{X}_n$ converges almost surely to μ):

$$P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$$

Strong LLN is a more powerful result (Strong LLN implies Weak LLN), but its proof is more complicated.

LLN - Example

How large a random sample must be taken from a given distribution in order for the probability to be at least 0.99 that the sample mean will be within 2 standard deviations of the mean of the distribution?

What about 0.95 probability to be within 1 standard deviations of the mean?

LLN and CLT

Law of large numbers shows us that

$$\lim_{n\to\infty} \frac{S_n - n\mu}{n} = \lim_{n\to\infty} (\bar{X}_n - \mu) \to 0$$

which shows that for large n, $n \gg \bar{S}_n - n\mu$.

What happens if we divide by something that grows slower than n like $\sqrt{n}$?

$$\lim_{n\to\infty} \frac{S_n - n\mu}{\sqrt{n}} = \lim_{n\to\infty} \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

This is the Central Limit Theorem, of which the DeMoivre-Laplace theorem for the normal approximation to the binomial is a special case. Hopefully by the end of this class we will have the tools to prove this.