Stat 343, Mon 21-Sep-2020 -- Mon 21-Sep-2020
Probability and Statistics
Fall 2020

-------------------------------
Monday, September 21st 2020
-------------------------------

Wk 4, Mo
Topic:: Independence
Topic:: Covariance

# Joint Distributions

Our calculation of probabilities has led to the consideration of the concurrence of two events—get a "spade" and a "king", roll "doubles" and a "number larger than 6", etc. And, as many events are depicted with random variables, this naturally leads to considering two or more random variables together. To facilitate answering questions such as $P(2 \leq X \leq 4 \text{ and } Y = 5)$, we would like to have (in the case where $X$, $Y$ are discrete r.v.s) a **joint pmf**, a function that yields values

$$f_{X,Y}(x,y) = P(X = x \,\text{and}\, Y = y) \qquad \text{abbreviated as } P(X = x, Y = y).$$

Naturally, the idea can be extended to that for a joint pmf of $k$ discrete r.v.s. If one has such a joint pmf, one easily recovers the individual (or **marginal**) distributions for $X$, $Y$:

$$f_X(x) = \sum_y f_{X,Y}(x,y) \qquad \text{and} \qquad f_Y(y) = \sum_x f_{X,Y}(x,y).$$

We can also obtain conditional distributions

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

(similar definition for $f_{Y|X=x}(y)$).

**Definition 1:** Suppose $f$ is the joint pmf of discrete r.v.s $X$, $Y$, and let $t\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Then $Z = t(X,Y)$ is a discrete r.v. with pmf given by

$$P(Z = z) \;=\; \sum_{\left\{(x,y) \,\middle|\, t(x,y)=z\right\}} f(x,y) \;=:\; \sum_{t(x,y)=z} f(x,y).$$

**Definition 2:** Suppose $f$ is the joint pmf of discrete r.v.s $X$, $Y$. We say $X$, $Y$ are **independent** if for every $x$ and $y$,

$$f(x,y) \;=\; f_X(x) \cdot f_Y(y).$$

**Theorem 1:** Let $X$, $Y$ be discrete random variables. Then

(i) $\mathrm{E}(X+Y) = \mathrm{E}(X) + \mathrm{E}(Y)$. [How about $\mathrm{E}(X-Y)$?]

(ii) $\mathrm{E}(XY) = \mathrm{E}(X) \cdot \mathrm{E}(Y)$, if $X$ and $Y$ are independent.

(iii) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$, if $X$ and $Y$ are independent.

Proof:

(i) Let $f(x,y)$ denote the joint pmf $\left(f_{XY}(x,y)\right)$.

$$E(X+Y) = \sum_x \sum_y (x+y) f(x,y) = \sum_x \sum_y \left[x f(x,y) + y f(x,y)\right]$$

$$= \sum_x \sum_y x f(x,y) + \sum_y \sum_x y f(x,y) = \sum_x x \sum_y f(x,y) + \sum_y y \sum_x f(x,y)$$

$$= \sum_x x \underbrace{f_X(x)}_{\text{marginal dist. for } X} + \sum_y y f_Y(y) = E(X) + E(Y).$$

(ii) 
$$E(XY) = \sum_{x,y} xy\, f(x,y) = \sum_x \sum_y xy\, f_X(x) f_Y(y)$$

$$= \sum_x \sum_y x f_X(x)\, y f_Y(y) = \sum_x x f_X(x) \underbrace{\left(\sum_y y f_Y(y)\right)}_{= E(Y)} = E(X)E(Y).$$

part (iii):
We have

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{E}\left((X+Y)^2\right) - [\mathrm{E}(X+Y)]^2 = \mathrm{E}\left(X^2+2XY+Y^2\right) - [\mathrm{E}(X+Y)]^2 \\
&= \mathrm{E}(X^2) + \mathrm{E}(2XY) + \mathrm{E}(Y^2) - [\mathrm{E}(X)]^2 - 2\,\mathrm{E}(X)\,\mathrm{E}(Y) - [\mathrm{E}(Y)]^2 \\
&= \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 + \mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2 + 2\,\mathrm{E}(XY) - 2\,\mathrm{E}(X)\,\mathrm{E}(Y) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2[\mathrm{E}(XY) - \mathrm{E}(X)\,\mathrm{E}(Y)] \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y). \qquad \text{(independence, part (ii))}
\end{aligned}$$

Define $E(XY) - E(X)E(Y) = \mathrm{Cov}(X,Y)$.

□



Note that $\mathrm{Cov}(X,X) = E(X \cdot X) - E(X) \cdot E(X) = \mathrm{Var}(X)$.

## Expected values and variances revisited

The results in the last theorem, when generalized to arbitrary sums of (independent) r.v.s, give us the tools for computing means and variances for several standard statistical models.

### Binomial distributions

**Special case**: $X \sim \mathsf{Binom}(1, \pi)$. Such an $X$ called a **Bernoulli random variable**. Here

$$\mu = \sum_x x f(x) = 0 \cdot (1-\pi) + 1 \cdot \pi = \pi$$
$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \pi - (\pi)^2 = \pi(1-\pi)$$

**General case**: $X \sim \mathsf{Binom}(n, \pi)$. Note

$$X = X_1 + X_2 + \cdots + X_n,$$

with each $X_j \sim \mathsf{Binom}(1, \pi)$ (Bernoulli) and the collection $X_1, \ldots, X_n$ is independent in the sense that

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(X_1 = x_1) \cdot P(X_2 = x_2) \cdots P(X_n = x_n).$$

By an analog to the last theorem,

$$\mathrm{E}(X) = E(X_1 + X_2 + \cdots + X_n) = \pi + \pi + \cdots + \pi = n\pi$$
$$\mathrm{Var}(X) = \mathrm{Var}(X_1 + X_2 + \cdots + X_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n)$$
$$= \pi(1-\pi) + \pi(1-\pi) + \cdots + \pi(1-\pi) = n\pi(1-\pi)$$

**Negative binomial distributions**.

In similar fashion, an $X \sim \mathsf{NBinom}(n, \pi)$ may be thought as the sum of independent random variables $X = X_1 + X_2 + \cdots + X_n$ where each $X_j \sim \mathsf{NBinom}(1, \pi)$. ($X_j$ counts the number of failed attempts between the $(j-1)^{\text{st}}$ success and the $j^{\text{th}}$ one.) We have not previously calculated the variance of a geometric r.v., but Pruim calculated the mean, on p. 79, to be $(1-\pi)/\pi$. Thus,

$$\mathrm{E}(X) = \sum_{j=1}^{n} \mathrm{E}(X_j) = \frac{1-\pi}{\pi} + \frac{1-\pi}{\pi} + \cdots + \frac{1-\pi}{\pi}$$
$$= \frac{n(1-\pi)}{\pi}$$

**The sum of independent, identically distributed r.v.s.** In both the binomial and negative binomial cases above, we could write $X = X_1 + X_2 + \cdots + X_n$, where all the $X_j$s come from the same distribution, and the collection $X_1, \ldots, X_n$ is independent. We abbreviate these assumptions about $X_1, \ldots, X_n$ by calling the i.i.d. random variables, where i.i.d. stands for *independent and identically distributed*. If $\mu, \sigma^2$ stand for the mean and variance, respectively, of the distribution common to the $X_j$, then their sum has mean and variance

$$\mathrm{E}(X) = \sum_{j=1}^{n} \mu = n\mu,$$

and

$$\mathrm{Var}(X) = \sum_{j=1}^{n} \sigma^2 = n\sigma^2.$$

standard dev. of $Y$ $= |a| \cdot$ (std. dev. of $X$)

**Lemma 1 (2.5.9):** Suppose $X$ is a discrete random variable, and $Y = aX + b$. Then

$$\mathrm{Var}(Y) = a^2\,\mathrm{Var}(X).$$

**Corollary 1:** Let $X_1, \ldots X_n$ represent an i.i.d. random sample from a population with mean $\mu$ and variance $\sigma^2$, and take the **sample mean** as the random variable

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Then

$$\mathrm{E}(X) = \mu, \qquad \text{and} \qquad \mathrm{Var}(X) = \frac{\sigma^2}{n}.$$

$$E(\overline{X}) = E\left(\frac{1}{n}(X_1 + \cdots + X_n)\right)$$
$$= \frac{1}{n} E(X_1 + \cdots + X_n) = \frac{1}{n}(\mu + \mu + \cdots \mu) = \mu$$

## Covariance

In the proof of the theorem concerning $\mathrm{Var}(X+Y)$, we arrived at

$$\mathrm{Var}(X+Y) \;=\; \mathrm{Var}(X) + \mathrm{Var}(Y) + 2[\mathrm{E}\,(XY) - \mathrm{E}\,(X)\,\mathrm{E}\,(Y)].$$

The quantity in brackets was zero in our proof because $X$, $Y$ were assumed to be independent. If $Y = X$, then they are certainly not independent, and the expression in the brackets is

$$\mathrm{E}\,(XX) - \mathrm{E}\,(X)\,\mathrm{E}\,(X) \;=\; \mathrm{Var}(X).$$

We define a new idea, the covariance of variables $X$, $Y$.

**Definition 3:** Let $X$, $Y$ be jointly distributed random variables. The **covariance** of $X$, $Y$ is

$$\mathrm{Cov}(X,Y) \;:=\; \mathrm{E}\,(XY) - \mathrm{E}\,(X)\,\mathrm{E}\,(Y).$$

**Lemma 2:** Let $X$, $Y$ be jointly distributed r.v.s. Then

$$\mathrm{Cov}(X,Y) \;=\; \mathrm{E}\,((X-\mu_X)(Y-\mu_Y)).$$

Argue for what can be known about covariance when $X$, $Y$ are

- simultaneously on the same side of their means.
- generally on opposite sides of their means.

**Lemma 3:** Let $X_1$, $X_2$, $\ldots$, $X_n$ be jointly distributed r.v.s. Then

$$\mathrm{Var}(X_1 + X_2 + \cdots + X_n) \;=\; \sum_i \mathrm{Var}(X_i) + 2\sum_{i<j}\mathrm{Cov}(X_i, X_j) \;=\; \sum_{i,j}\mathrm{Cov}(X_i, X_j).$$

**Definition 4:** The **correlation coefficient** $\rho$ of random variables $X$, $Y$ is defined to be

$$\rho \;:=\; \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}} \;=:\; \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

**Lemma 4:** The correlation coefficient $\rho$ of r.v.s $X$, $Y$ lies in the interval $[-1, 1]$.

Proof: See the text, p. 91. $\square$

**Lemma 5:** Suppose $X$, $Y$ are jointly distributed r.v.s such that $\rho = \pm 1$. Then there are constants $a$, $b$ such that $P(Y = a + bX) = 1$.

Proof: See the text, p. 91. Note: When $\rho = -1$, the slope $b$ is negative; when $\rho = 1$, the slope $b$ is positive. $\square$