Handout 8: Math/Stat 394: Probability I

Expectation, Mean, Standard, and Variance

Wellner; 2/16/2000

Definitions:

$$
\begin{aligned}
Eg(X) &= \text{the expectation of } \ g(X) \\
&= \begin{cases} \sum_x g(x)p(x), & \text{in the discrete case} \\ \int g(x)f(x)dx, & \text{in the continuous case}\,; \end{cases} \\
\mu = E(X) = EX &= \text{the true mean of } \ X \\
&= \begin{cases} \sum_x xp(x), & \text{in the discrete case} \\ \int xf(x)dx, & \text{in the continuous case}\,; \end{cases} \\
\tau = E|X-\mu| &= \text{the mean deviation of } \ X (\text{about its mean}) \\
&= \begin{cases} \sum_x |x-\mu|p(x), & \text{in the discrete case} \\ \int |x-\mu|f(x)dx, & \text{in the continuous case}\,; \end{cases} \\
\sigma^2 \equiv Var(X) \equiv E(X-\mu)^2 &= \text{the variance of } \ X \\
&= \begin{cases} \sum_x (x-\mu)^2 p(x), & \text{in the discrete case} \\ \int (x-\mu)^2 f(x)dx, & \text{in the continuous case}\,; \end{cases} \\
\sigma = \sqrt{\sigma^2} = \sqrt{Var(X)} &= \ \text{the standard deviation of } \ X\,; \\
Cov(X,Y) \equiv E(X-EX)(Y-EY) &= \ \text{the covariance between } X \text{ and } Y\,; \\
\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} &= \ \text{the correlation between } X \text{ and } Y\,.
\end{aligned}
$$

Example 1. Here are four discrete distributions on $\{1, \ldots, 6\}$. The third distribution is the "most spread out" or "most variable", while the fourth distribution is th "least spread out" or "least variable". (You should draw the corresponding pictures of the probability mass functions!)

1	2	3	4	5	6	μ	σ	τ
1/6	1/6	1/6	1/6	1/6	1/6	3.5	1.708	1.5
1/21	2/21	3/21	4/21	5/21	6/21	4.333	1.491	1.2698
6/21	1/21	2/21	3/21	4/21	5/21	3.619	1.963	1.769
1/12	2/12	3/12	3/12	2/12	1/12	3.5	1.384	1.166

Example 2. If $X \sim$ Bernoulli(p), then
$E(X) = 1 \cdot p + 0 \cdot q = p$,
$Var(X) = E(X - \mu)^2 = (1 - p)^2 \cdot p + (0 - p)^2 \cdot q = pq$,
$\sigma = \sqrt{Var(X)} = \sqrt{pq}$,
and $\tau = E|X - \mu| = |1 - p| \cdot p + |0 - p| \cdot q = 2pq$. Note that $\tau = 2pq \leq \sqrt{pq} = \sigma$ always with equality iff $p \in \{0, 1/2, 1\}$.

Example 3. Here is one more discrete distribution with a table to illustrate the calculation of μ, τ, and σ:

x	$p(x)$	$xp(x)$	$\lvert x - \mu\rvert p(x)$	$\lvert x - \mu\rvert p(x)$	$(x - \mu)^2$	$(x - \mu)^2 p(x)$
1	.2	.2	1.4	.28	1.96	.392
2	.3	.6	.4	.12	.16	.096
3	.4	1.2	.6	.24	.36	.432
4	.1	.4	1.6	.16	2.56	1.024
	1	2.4		.80		1.944

Thus we see that for this distribution $\mu = 2.4$, $\tau = .8$, $\sigma^2 = 1.944$, and $\sigma = 1.394$.

Example 4. For the triangular distribution which appeared as the marginal distribution for the Example in HO #7, namely $f_X(x) = 2x1_{(0,1)}(x)$, we have

$$\mu = E(X) = \int x f_X(x) dx = \int_0^1 x 2x dx = \frac{2}{3},$$

$$Var(X) = E(X-\mu)^2 = E(X^2)-\mu^2 = \int_0^1 x^2 2x dx-(2/3)^2 = \frac{2}{4}-(2/3)^2 = \frac{1}{18}\,,$$

$$\sigma = \sqrt{Var(X)} = \sqrt{1/18} = 0.2357\,,$$

and

$$\begin{aligned}
\tau &= E|X-\mu| = \int_0^1 |x-2/3| 2x dx \\
&= \int_0^{2/3} (2/3-x)2x dx + \int_{2/3}^1 (x-2/3)2x dx \\
&= \left(\frac{4}{3}\frac{x^2}{2} - \frac{2}{3}x^3 \,\Big|_0^{2/3} + \left(\frac{2}{3}x^3 - \frac{4}{3}\frac{x^2}{2}\,\Big|_{2/3}^1 \right.\right. \\
&= \left(\frac{2}{3}\right)^3 - \left(\frac{2}{3}\right)^4 + \left(\frac{2}{3} - \frac{2}{3}\right) - \left(\left(\frac{2}{3}\right)^4 - \left(\frac{2}{3}\right)^3\right) \\
&= 2\left\{\left(\frac{2}{3}\right)^3 - \left(\frac{2}{3}\right)^4\right\} \\
&= \left(\frac{2}{3}\right)^4 \doteq .1975\ldots < .2357 = \sigma\,.
\end{aligned}$$

Similarly, for the marginal distribution of Y, with density $f_Y(y) = 2(1-y)1_{(0,1)}(y)$,

$$\mu_Y = E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_0^1 y2(1-y)dy = \frac{1}{3}\,, \tag{0.1}$$

$$Var(Y) = E(Y-E(Y))^2 = E(Y^2) - (EY)^2 = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}\,,$$

and

$$\tau_Y = E|Y - EY| = (2/3)^4\,.$$

(These are all easy once we note that Y has the same distribution as $1-X$, and then use the Facts which we develop below!) For the bivariate density which is uniform on the triangle T given in the Example of HO # 7, we compute

$$\begin{aligned}
E(XY) &= \int_0^1 \int_0^1 xy2 \cdot 1_{[y\le x]} dy dx \\
&= \int_0^1 2x \left(\int_0^x y dy\right) dx = \int_0^1 x^3 dx = \frac{1}{4}\,,
\end{aligned}$$

and hence

$$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{4} - \frac{2}{3}\cdot\frac{1}{3} = \frac{1}{36}\,.$$

Thus we also compute

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{1/36}{\sqrt{(1/18)(1/18)}} = \frac{1}{2}\,.$$

Now we proceed to develop some basic facts about expectation:

Fact 1: $E(aX+b) = aE(X)+b$.

Fact 2: $E(aX+bY+c) = aE(X)+bE(Y)+c$.

Fact 3: $Var(aX+b) = a^2Var(X)$.

Fact 4: $Var(aX+bY+c) = a^2Var(X) + 2abCov(X,Y) + b^2Var(Y)$.

Fact 5: $Cov(aX+c, bY+d) = abCov(X,Y)$.

Fact 6: $\rho_{aX+c,bY+d} = \text{sign}(ab)\rho_{X,Y}$. Here $\text{sign}(c) = 1$ if $c \geq 0$, $\text{sign}(c) = -1$ if $c < 0$.

Computing Formulas:
$Var(X) = E(X^2) - (EX)^2$;
$Cov(X,Y) = E(XY) - EX\cdot EY$;
$Var(X) = Cov(X,X)$;
$Cov(X,Y) = 0$ whenever X and Y are independent.

Proof of Fact 1: In the discrete case, with $p(x) \equiv P(X=x)$,

$$\begin{aligned} E(aX+b) &= \sum_x (ax+b)p(x) \\ &= a\sum_x xp(x) + b\sum_x p(x) \\ &= aE(X) + b\cdot 1 = aE(X)+b\,. \end{aligned}$$

In the continuous case,

$$\begin{aligned} E(aX+b) &= \int (ax+b)f(x)dx \\ &= a\int xf(x)dx + b\int f(x)dx \\ &= aE(X) + b\cdot 1 = aE(X)+b\,. \end{aligned}$$

Proof of Fact 2: In the discrete case, with $p(x,y) \equiv P(X=x, Y=y)$, $p_X(x) = P(X=x)$, and $p_Y(y) = P(Y=y)$,

$$
\begin{aligned}
E(aX + bY + c) &= \sum_x (ax + by + c)p(x,y) \\
&= a\sum_{x,y} xp(x,y) + b\sum_{x,y} yp(x,y) + c\sum_{x,y} p(x,y) \\
&= a\sum_x xp_X(x) + b\sum_y yp_Y(y) + c \cdot 1 = aE(X) + bE(Y) + c\,.
\end{aligned}
$$

In the continuous case, with joint density $f(x,y)$ with marginal densities f_X and f_Y of X and Y,

$$
\begin{aligned}
E(aX + bY + c) &= \int\int (ax + by + c)f(x,y)dxdy \\
&= a\int\int xf(x,y)dxdy + b\int\int yf(x,y)dxdy + c\int\int f(x,y)dxdy \\
&= a\int xf_X(x)dx + b\int yf_Y(y)dy + c \cdot 1 \\
&= aE(X) + b \cdot E(Y) + c\,.
\end{aligned}
$$

Proof of Fact 3: Note that in either the continuous or discrete case,

$$
\begin{aligned}
Var(aX + b) &= E\{aX + b - E(aX + b)\}^2 \\
&= E\{aX + b - aE(X) - b\}^2 \qquad \text{using Fact 1} \\
&= E\{a(X - E(X))\}^2 \qquad \text{by algebra} \\
&= E\{a^2(X - E(X))^2\} \\
&= a^2E(X - E(X))^2 \qquad \text{by Fact 1} \\
&= a^2Var(X)\,.
\end{aligned}
$$

Note that this yields $\sigma_{aX} = |a|\sigma_X$.

Proof of Fact 5: By the definition of Covariance,

$$
Cov(aX + c, bY + d) = E[(aX + c - E(aX + c))(bY + d - E(bY + d))]
$$

$$
\begin{aligned}
&= E[(aX + c - aE(X) - c)(bY + d - bE(Y) - d)] \qquad \text{by Fact 1 twice} \\
&= E[(a(X - EX))(b(Y - EY))] \qquad \text{by algebra} \\
&= E[ab(X - EX)(Y - EY)] \\
&= abE[(X - EX)(Y - EY)] \qquad \text{by Fact 1} \\
&= abCov(X, Y) \qquad \text{by definition of Cov} .
\end{aligned}
$$

Proof of Fact 4: First we prove a simpler version, namely

$$Var(X + Y) = Var(X) + 2Cov(X, Y) + Var(Y) .$$

I'll call this **Fact** $4'$**.** To see this, note that

$$
\begin{aligned}
Var(X + Y) &= E(X + Y - E(X + Y))^2 \\
&= E(X + Y - E(X) - E(Y))^2 \qquad \text{by Fact 2} \\
&= E\{[(X - EX) + (Y - EY)]^2\} \qquad \text{by regrouping terms} \\
&= E\{(X - EX)^2 + 2(X - EX)(Y - EY) + (Y - EY)^2\} \qquad \text{by algebra} \\
&= E(X - EX)^2 + 2E(X - EX)(Y - EY) + E(Y - EY)^2 \qquad \text{by Fact 2} \\
&= Var(X) + 2Cov(X, Y) + Var(Y) \qquad \text{by definition of Var, Cov} .
\end{aligned}
$$

Now to prove Fact 4 itself, we write

$$
\begin{aligned}
Var(aX + bY + c) &= Var(aX + bY) \qquad \text{by Fact 3} \\
&= Var(aX) + 2Cov(aX, bY) + Var(bY) \qquad \text{by Fact } 4' \\
&= a^2Var(X) + 2abCov(X, Y) + b^2Var(Y) \qquad \text{by Facts 3 and 5} .
\end{aligned}
$$

Example 5. Uniform distributions. Suppose that $U \sim \text{Uniform}(0, 1)$. Then

$$\mu = E(U) = \int_0^1 x\,dx = \frac{1}{2}x^2|_0^1 = \frac{1}{2} ;$$

$$\sigma^2 = E(X^2) - \mu^2 = \int_0^1 x^2 dx - \mu^2 = (1/3) - (1/4) = 1/12 ;$$

so $\sigma = \sqrt{1/12} \doteq .2887$. The mean deviation is

$$
\begin{aligned}
\tau = E|U - 1/2| &= \int_0^1 |x - 1/2|dx = 2\int_{1/2}^1 (x - 1/2)dx \\
&= (x - 1/2)^2|_{1/2}^1 = 1/4 < .2887 \doteq \sigma .
\end{aligned}
$$

Furthermore, note that the distribution function (cdf) for U is given by

$$F(x) = P(U \leq x) = \int_{-\infty}^{x} f(y)dy = \int_{0}^{x} 1 \cdot dy = x$$

if $0 \leq x \leq 1$, while it equals 0 if $x \leq 0$ and 1 if $x \geq 1$.

Example 6. For $U \sim$ Uniform$(0,1)$ as in Example 5, and real numbers $a < b$, set $X = (b-a)U + a$. Note that X takes values in $[a,b]$; in fact, the distribution of X is Uniform(a,b): $f_X(x) = (b-a)^{-1} 1_{[a,b]}(x)$. Now from Fact 1,

$$E(X) = E((b-a)U + a) = (b-a)E(U) + a = (b-a)/2 + a = (b+a)/2\,,$$

and from Fact 3,

$$Var(X) = Var((b-a)U+a) = (b-a)^2 Var(U) = (b-a)^2/12\,; \qquad \sigma_X = (b-a)/\sqrt{12}\,.$$

Example 7. Suppose $Y \sim$ Exponential(ν); i.e.

$$f_Y(y) = \nu e^{-\nu y} 1_{[0,\infty)}(y)\,.$$

Now

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = \int_{-\infty}^{y} f_Y(y')dy' \\ &= \int_{0}^{y} \nu e^{-\nu y'} dy' = -e^{-\nu y'}\big|_0^y \\ &= 1 - e^{-\nu y} \qquad \text{for} \quad y \geq 0\,. \end{aligned}$$

Thus $P(Y > y) = e^{-\nu y}$ for $y \geq 0$; this is the "survival probability". Now

$$\begin{aligned} E(Y) &= \int_{0}^{\infty} y f_Y(y)dy = \int_{0}^{\infty} y\nu e^{-\nu y} dy \\ &= \nu^{-1} \int_{0}^{\infty} te^{-t}dt \quad \text{by letting } t = \nu y \\ &= \nu^{-1}\,. \end{aligned}$$

Similarly

$$\begin{aligned} E(Y^2) &= \int_0^\infty y^2 f_Y(y)dy = \int_0^\infty y^2 \nu e^{-\nu y} dy \\ &= \nu^{-2} \int_0^\infty t^2 e^{-t} dt \quad \text{by letting } t = \nu y \\ &= 2\nu^{-2}. \end{aligned}$$

Thus

$$\sigma^2 = E(Y - E(Y))^2 = E(Y^2) - \mu^2 = 2\nu^{-2} - \nu^{-2} = 1/\nu^2 .$$

and $\sigma = 1/\nu$.
Fact: For $a > 0$, $F_{aY}(t) = P(aY \le t) = P(Y \le t/a) = F_Y(t/a)$.
Thus $f_{aY}(t) = (1/a) f_Y(t/a) = (\nu/a)e^{-(\nu/a)t}$ for $t \ge 0$; i.e. $aY \sim$ Exponential(ν/a).
Note that

$$P(Y > s + t | Y > s) = \frac{P(Y > s+t)}{P(Y > s)} = \frac{e^{-\nu(s+t)}}{e^{-\nu s}} = e^{-\nu t} = P(Y > t) .$$

This is the *lack of memory* property of the exponential distribution. Moreover,

$$\begin{aligned} \lambda(y) &\equiv \frac{f_Y(y)}{1 - F_Y(y)} = \text{(the instantaneous failure rate)} \\ &= \frac{\nu e^{-\nu y}}{e^{-\nu y}} = \nu, \quad \text{a constant} . \end{aligned}$$