Today.

Variance, covariance.

Discuss expectation as predictor.

- How close to expectation? Using expectation, and variance.
- What if you predict expectation?

Also, prediction from evidence.

Calculating $E[g(X)]$: LOTUS

Let $Y = g(X)$. Assume that we know the distribution of X.

We want to calculate $E[Y]$.

Method 1: We calculate the distribution of Y:

$$Pr[Y = y] = Pr[X \in g^{-1}(y)] \text{ where } g^{-1}(x) = \{x \in \Re : g(x) = y\}.$$

This is typically rather tedious!

Method 2: We use the following result.

Called "Law of the unconscious statistician."

Theorem:

$$E[g(X)] = \sum_x g(x) Pr[X = x].$$

Proof:

$$\begin{aligned}
E[g(X)] &= \sum_\omega g(X(\omega)) Pr[\omega] = \sum_x \sum_{\omega \in X^{-1}(x)} g(X(\omega)) Pr[\omega] \\
&= \sum_x \sum_{\omega \in X^{-1}(x)} g(x) Pr[\omega] = \sum_x g(x) \sum_{\omega \in X^{-1}(x)} Pr[\omega] \\
&= \sum_x g(x) Pr[X = x].
\end{aligned}$$

$\square$

Poll.

Which is LOTUS?

(A) $E[X] = \sum_{x \in \text{Range}(X)} g(x) \times Pr[g(X) = g(x)]$
No. overcounts $Pr[g(X) = g(x)]$.

(B) $E[X] = \sum_{x \in \text{Range}(X)} g(x) \times Pr[X = x]$
Yes. May count $g(x)$ twice, if $g(x) = g(x')$.

(C) $E[X] = \sum_{x \in \text{Range}(g)} x \times Pr[g(X) = x]$
No. $g(x)$ is image, x is pre-image.

Geometric Distribution.

Experiment: flip a coin with heads prob. p. until Heads.
Random Variable X: number of flips.

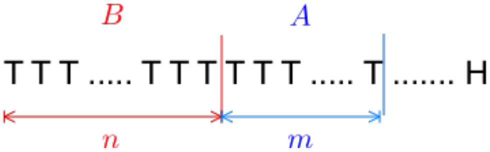
And distribution is:

(A) $X \sim G(p) : Pr[X = i] = (1-p)^{i-1}p$.
(B) $X \sim B(p,n) : Pr[X = i] = \binom{n}{i} p^i (1-p)^{n-i}$.

(A) Distribution of $X \sim G(p)$: $Pr[X = i] = (1-p)^{i-1}p$.

Geometric Distribution: Memoryless - Interpretation

$$Pr[X > n+m | X > n] = Pr[X > m], m, n \geq 0.$$



$$Pr[X > n+m | X > n] = Pr[A|B] = Pr[A'] = Pr[X > m].$$

A': is m coin tosses before heads.
$A|B$: m 'more' coin tosses before heads.

The coin is memoryless , therefore, so is X.
Independent coin: $Pr[H|\text{'any previous set of coin tosses'}] = p$

Geometric Distribution: Memoryless by derivation.

Let X be $G(p)$. Then, for $n \geq 0$,

$$Pr[X > n] = Pr[\text{ first } n \text{ flips are } T] = (1-p)^n.$$

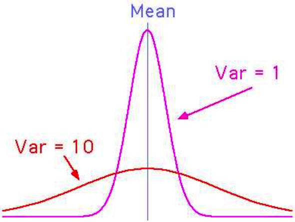
Theorem

$$Pr[X > n+m | X > n] = Pr[X > m], m, n \geq 0.$$

Proof:

$$\begin{aligned} Pr[X > n+m | X > n] &= \frac{Pr[X > n+m \text{ and } X > n]}{Pr[X > n]} \\ &= \frac{Pr[X > n+m]}{Pr[X > n]} \\ &= \frac{(1-p)^{n+m}}{(1-p)^n} = (1-p)^m \\ &= Pr[X > m]. \end{aligned}$$

Variance



The variance measures the deviation from the mean value.

Definition: The variance of X is

$$\sigma^2(X) := var[X] = E[(X - E[X])^2].$$

$\sigma(X)$ is called the standard deviation of X.

Variance and Standard Deviation

Fact:

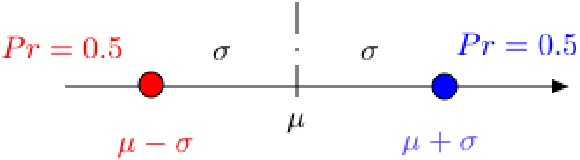
$$var[X] = E[X^2] - E[X]^2.$$

Indeed:

$$\begin{aligned} var(X) &= E[(X - E[X])^2] \\ &= E[X^2 - 2XE[X] + E[X]^2) \\ &= E[X^2] - 2E[X]E[X] + E[X]^2, \text{ by linearity} \\ &= E[X^2] - E[X]^2. \end{aligned}$$

A simple example

This example illustrates the term 'standard deviation.'



Consider the random variable X such that

$$X = \begin{cases} \mu - \sigma, & \text{w.p. } 1/2 \\ \mu + \sigma, & \text{w.p. } 1/2. \end{cases}$$

Then, $E[X] = \mu$ and $(X - E[X])^2 = \sigma^2$. Hence,

$$var(X) = \sigma^2 \text{ and } \sigma(X) = \sigma.$$

Example

Consider X with

$$X = \begin{cases} -1, & \text{w. p. } 0.99 \\ 99, & \text{w. p. } 0.01. \end{cases}$$

Then

$$\begin{aligned} E[X] &= -1 \times 0.99 + 99 \times 0.01 = 0. \\ E[X^2] &= 1 \times 0.99 + (99)^2 \times 0.01 \approx 100. \\ Var(X) &\approx 100 \implies \sigma(X) \approx 10. \end{aligned}$$

Also,

$$E(|X|) = 1 \times 0.99 + 99 \times 0.01 = 1.98.$$

Thus, $\sigma(X) = \sqrt{E[(X - E(X))^2]} \neq E[|X - E[X]|]$!

Exercise: How big can you make $\frac{\sigma(X)}{E[|X-E[X]|]}$?

Roughly square root of max value, M.

Keep expectation small using $1/M$.

Yields $2E[X]$ for $E[|X - E[X]|]$, and $\approx \sqrt{M}$ for $\approx E[\sqrt{X - E(X)}]$.

Uniform

Assume that $Pr[X = i] = \frac{1}{n}$ for $i \in \{1, \ldots, n\}$. Then

$$\begin{aligned} E[X] &= \sum_{i=1}^{n} i \times Pr[X = i] = \frac{1}{n} \sum_{i=1}^{n} i \\ &= \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}. \end{aligned}$$

Also,

$$\begin{aligned} E[X^2] &= \sum_{i=1}^{n} i^2 \times Pr[X = i] = \frac{1}{n} \sum_{i=1}^{n} i^2 \\ &= \frac{1}{n} \frac{(n)(n+1)(n+2)}{6} = \frac{1+3n+2n^2}{6}, \text{ as you can verify.} \end{aligned}$$

This gives

$$var(X) = \frac{1+3n+2n^2}{6} - \frac{(n+1)^2}{4} = \frac{n^2-1}{12}.$$

(Sort of $\int_0^{1/2} x^2 dx = \frac{x^3}{3}$.)

Variance of geometric distribution.

X is a geometrically distributed RV with parameter p.
Thus, $Pr[X=n]=(1-p)^{n-1}p$ for $n \geq 1$. Recall $E[X]=1/p$.

$$\begin{aligned}
E[X^2] &= p+4p(1-p)+9p(1-p)^2+... \\
-(1-p)E[X^2] &= \quad -[p(1-p)+4p(1-p)^2+...] \\
pE[X^2] &= p+3p(1-p)+5p(1-p)^2+... \\
&= 2(p+2p(1-p)+3p(1-p)^2+..) \quad E[X]! \\
& \quad -(p+p(1-p)+p(1-p)^2+...) \quad \text{Distribution.} \\
pE[X^2] &= 2E[X]-1 \\
&= 2(\frac{1}{p})-1=\frac{2-p}{p}
\end{aligned}$$

$\implies E[X^2]=(2-p)/p^2$ and
$var[X]=E[X^2]-E[X]^2=\frac{2-p}{p^2}-\frac{1}{p^2}=\frac{1-p}{p^2}$.
$\sigma(X)=\frac{\sqrt{1-p}}{p} \approx E[X]$ when p is small(ish).

Fixed points.

Number of fixed points in a random permutation of n items.
"Number of student that get homework back."

$X = X_1 + X_2 \cdots + X_n$

where X_i is indicator variable for ith student getting hw back.

$$\begin{aligned} E(X^2) &= \sum_i E(X_i^2) + \sum_{i \neq j} E(X_i X_j). \\ &= n \times \frac{1}{n} + (n)(n-1) \times \frac{1}{n(n-1)} \\ &= 1 + 1 = 2. \end{aligned}$$

$E(X_i^2) = 1 \times Pr[X_i = 1] + 0 \times Pr[X_i = 0]$
$= \frac{1}{n}$
$E(X_i X_j) = 1 \times Pr[X_i = 1 \cap X_j = 1] + 0 \times Pr[\text{"anything else"}]$
$= 1 \times \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$

$Var(X) = E(X^2) - (E(X))^2 = 2 - 1 = 1.$

Poll: fixed points.

What's true?

(A) X_i and X_j are independent.
No. If student i gets student j's homework.

(B) $E[X_i X_j] = Pr[X_i X_j = 1]$
Yes. Indicator random variable.

(C) $Pr[X_i X_j] = \frac{(n-2)!}{n!}$
Yes. $(n-2)!$ outcomes where $X_i X_j = 1$.

(D) $X_i^2 = X_i$.
Yes. $1^2 = 1$ and $0^2 = 1$, $X_i \in \{0, 1\}$.

Variance: binomial.

$$\begin{aligned} E[X^2] &= \sum_{i=0}^{n} i^2 \times \binom{n}{i} p^i (1-p)^{n-i}. \\ &= \text{Really???!!\#\#...} \end{aligned}$$

Too hard!

Ok.. fine.
Let's do something else.
Maybe not much easier...but there is a payoff.

Properties of variance.

1. $Var(cX) = c^2 Var(X)$, where c is a constant.
   Scales by c^2.
2. $Var(X + c) = Var(X)$, where c is a constant.
   Shifts center.

Proof:

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

□

Independent random variables.

Independent: $P[X=a, Y=b] = Pr[X=a]Pr[Y=b]$

Fact: $E[XY] = E[X]E[Y]$ for independent random variables.

$$\begin{aligned}
E[XY] &= \sum_a \sum_b a \times b \times Pr[X=a, Y=b] \\
&= \sum_a \sum_b a \times b \times Pr[X=a]Pr[Y=b] \\
&= (\sum_a aPr[X=a])(\sum_b bPr[Y=b]) \\
&= E[X]E[Y]
\end{aligned}$$

Variance of sum of two independent random variables

Theorem:
If X and Y are independent, then

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

Proof:
Since shifting the random variables does not change their variance, let us subtract their means.

That is, we assume that $E(X) = 0$ and $E(Y) = 0$.

Then, by independence,

$$E(XY) = E(X)E(Y) = 0.$$

Hence,

$$\begin{aligned} var(X+Y) &= E((X+Y)^2) = E(X^2 + 2XY + Y^2) \\ &= E(X^2) + 2E(XY) + E(Y^2) = E(X^2) + E(Y^2) \\ &= var(X) + var(Y). \end{aligned}$$

Variance of sum of independent random variables

Theorem:
If $X, Y, Z, \dots$ are pairwise independent, then

$$var(X + Y + Z + \cdots) = var(X) + var(Y) + var(Z) + \cdots .$$

Proof:
Since shifting the random variables does not change their variance, let us subtract their means.

That is, we assume that $E[X] = E[Y] = \cdots = 0$.

Then, by independence,

$$E[XY] = E[X]E[Y] = 0. \text{ Also, } E[XZ] = E[YZ] = \cdots = 0.$$

Hence,

$$\begin{aligned} var(X + Y + Z + \cdots) &= E((X + Y + Z + \cdots)^2) \\ &= E(X^2 + Y^2 + Z^2 + \cdots + 2XY + 2XZ + 2YZ + \cdots) \\ &= E(X^2) + E(Y^2) + E(Z^2) + \cdots + 0 + \cdots + 0 \\ &= var(X) + var(Y) + var(Z) + \cdots . \end{aligned}$$

$\square$

Variance of Binomial Distribution.

Flip coin with heads probability p.
X- how many heads?

$$X_i = \begin{cases} 1 & \text{if } i\text{th flip is heads} \\ 0 & \text{otherwise} \end{cases}$$

$E(X_i^2) = 1^2 \times p + 0^2 \times (1-p) = p$.
$Var(X_i) = p - (E(X))^2 = p - p^2 = p(1-p)$.

$p = 0 \implies Var(X_i) = 0$
$p = 1 \implies Var(X_i) = 0$

$X = X_1 + X_2 + \ldots X_n$.

X_i and X_j are independent: $Pr[X_i = 1 | X_j = 1] = Pr[X_i = 1]$.

$$Var(X) = Var(X_1 + \cdots X_n) = np(1-p).$$

Poisson Distribution: Variance.

Definition Poisson Distribution with parameter $\lambda > 0$

$$X = P(\lambda) \Leftrightarrow Pr[X = m] = \frac{\lambda^m}{m!}e^{-\lambda}, m \geq 0.$$

Mean, Variance?

Ugh.

Recall that Poission is the limit of the Binomial with $p = \lambda/n$ as $n \to \infty$.

Mean: $pn = \lambda$

Variance: $p(1-p)n = \lambda - \lambda^2/n \to \lambda$.

$E(X^2)$? $Var(X) = E(X^2) - (E(X))^2$ or $E(X^2) = Var(X) + E(X)^2$.

$E(X^2) = \lambda + \lambda^2$.

Covariance

Definition The covariance of X and Y is

$$cov(X, Y) := E[(X - E[X])(Y - E[Y])].$$

Fact

$$cov(X, Y) = E[XY] - E[X]E[Y].$$

Proof:
Think about $E[X] = E[Y] = 0$. Just $E[XY]$. $\square$ish.

For the sake of completeness.

$$E[(X - E[X])(Y - E[Y])] = E[XY - E[X]Y - XE[Y] + E[X]E[Y]]$$
$$= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y]$$
$$= E[XY] - E[X]E[Y].$$

Correlation

Definition The correlation of X, Y, $Cor(X,Y)$ is

$$corr(X,Y) : \frac{cov(X,Y)}{\sigma(X)\sigma(Y)}.$$

Theorem: $-1 \leq corr(X,Y) \leq 1$.
Proof: Idea: $(a-b)^2 > 0 \rightarrow a^2 + b^2 \geq 2ab$.

Simple case: $E[X] = E[Y] = 0$ and $E[X^2] = E[Y^2] = 1$.

$Cov(X,Y) = E[XY]$.

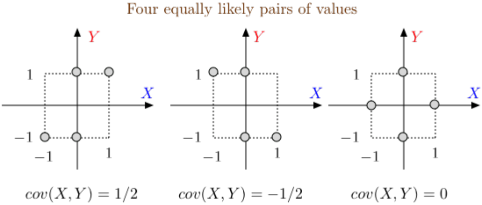
$E[(X-Y)^2] = E[X^2] + E[Y^2] - 2E[XY] = 2(1 - E[XY]) \geq 0$
$\rightarrow E[XY] \leq 1$.

$E[(X+Y)^2] = E[X^2] + E[Y^2] + 2E[XY] = 2(1 + E[XY]) \geq 0$
$\rightarrow E[XY] \geq -1$.

□

Shifting and scaling doesn't change correlation.

Examples of Covariance



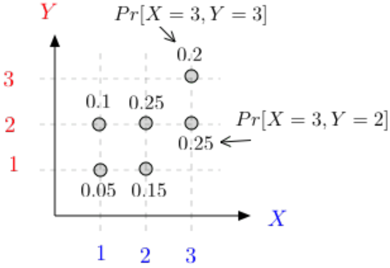
Note that $E[X] = 0$ and $E[Y] = 0$ in these examples. Then $cov(X,Y) = E[XY]$.

When $cov(X,Y) > 0$, the RVs X and Y tend to be large or small together. X and Y are said to be positively correlated.

When $cov(X,Y) < 0$, when X is larger, Y tends to be smaller. X and Y are said to be negatively correlated.

When $cov(X,Y) = 0$, we say that X and Y are uncorrelated.

Examples of Covariance



$$E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 2.3$$
$$E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8$$
$$E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2$$
$$E[Y^2] = 1 \times 0.2 + 4 \times 0.6 + 9 \times 0.2 = 4.4$$
$$E[XY] = 1 \times 0.05 + 1 \times 2 \times 0.1 + \cdots + 3 \times 3 \times 0.2 = 4.85$$
$$cov(X, Y) = E[XY] - E[X]E[Y] = .25$$
$$var[X] = E[X^2] - E[X]^2 = .51$$
$$var[Y] = E[Y^2] - E[Y]^2 = .4$$
$$corr(X, Y) \approx 0.55$$

Properties of Covariance

$$cov(X,Y) = E[(X-E[X])(Y-E[Y])] = E[XY] - E[X]E[Y].$$

Fact

(a) $var[X] = cov(X,X)$

(b) X, Y independent $\Rightarrow cov(X,Y) = 0$

(c) $cov(a+X, b+Y) = cov(X,Y)$

(d) $cov(aX+bY, cU+dV) = ac \cdot cov(X,U) + ad \cdot cov(X,V) + bc \cdot cov(Y,U) + bd \cdot cov(Y,V).$

Proof:

(a)-(b)-(c) are obvious.

(d) In view of (c), one can subtract the means and assume that the RVs are zero-mean. Then,

$$\begin{aligned} cov(aX+bY, cU+dV) &= E[(aX+bY)(cU+dV)] \\ &= ac \cdot E[XU] + ad \cdot E[XV] + bc \cdot E[YU] + bd \cdot E[YV] \\ &= ac \cdot cov(X,U) + ad \cdot cov(X,V) + bc \cdot cov(Y,U) + bd \cdot cov(Y,V). \end{aligned}$$

□

Lake Woebegone: Poll

What is true?

(A) Everyone is above average (on midterm)
False. Average would be higher.

(B) For a random variable, at most half the people can be more than twice the average.
False. Consder $Pr[X = -2] = 1/3$ and $Pr[X = 1] = 2/3$. $E[X] = 0$.

(C) For the midterm with no negative scores, at most half the people can be more than twice the average.
True. Otherwise average would be higher.

Markov's inequality

The inequality is named for Andrey Markov, though in work by Pafnuty Chebyshev. (Sometimes) called Chebyshev's first inequality.

Theorem Markov's Inequality
Assume $f : \Re \rightarrow [0, \infty)$ is nondecreasing. Then,

$$Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}, \text{ for all } a \text{ such that } f(a) > 0.$$

Proof:

Claim:

$$1\{X \geq a\} \leq \frac{f(X)}{f(a)}.$$

If $X < a$, the inequality reads $0 \leq f(x)/f(a)$, since $f(\cdot) \geq 0$.
If $X \geq a$, it reads $1 \leq f(x)/f(a)$, since $f(\cdot)$ is nondecreasing.

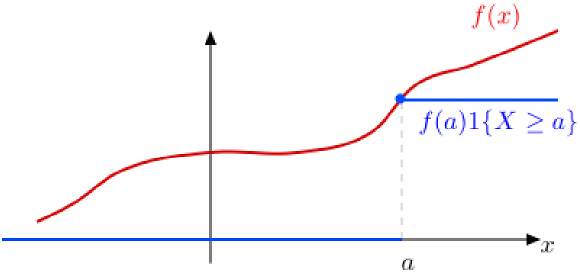
Taking the expectation yields the inequality,
expectation of an indicator is the probability.
and expectation is monotone, e.g., weighted sum of points.

That is, $\sum_v Pr[X = v] 1\{v \geq a\} \leq \sum_v Pr[X = v] \frac{f(v)}{f(a)}$.

□

Intuition: $E[f(X)] \geq f(a) Pr[X > a] = f(a) Pr[X > f(a)]$.

A picture



$$f(a)1\{X \geq a\} \leq f(x) \Rightarrow 1\{X \geq a\} \leq \frac{f(X)}{f(a)}$$

$$\Rightarrow Pr[X \geq a] \leq \frac{E[f(X)]}{f(a)}$$

Markov Inequality Example: G(p)

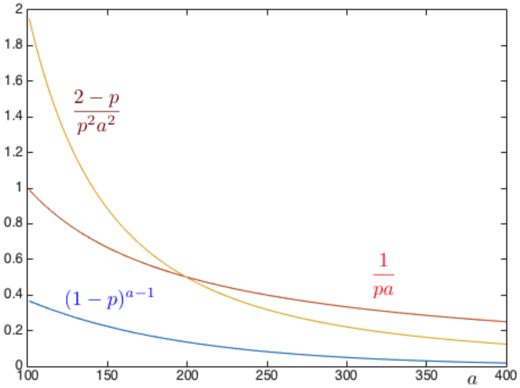
Let $X = G(p)$. Recall that $E[X] = \frac{1}{p}$ and $E[X^2] = \frac{2-p}{p^2}$.

Choosing $f(x) = x$, we get

$$Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{1}{ap}.$$

Choosing $f(x) = x^2$, we get

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{2-p}{p^2 a^2}.$$



Markov Inequality Example: $P(\lambda)$

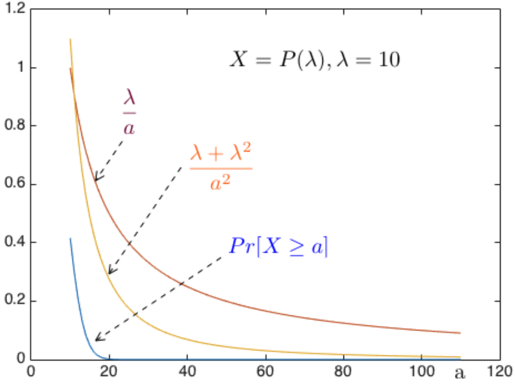
Let $X = P(\lambda)$. Recall that $E[X] = \lambda$ and $E[X^2] = \lambda + \lambda^2$.

Choosing $f(x) = x$, we get

$$Pr[X \geq a] \leq \frac{E[X]}{a} = \frac{\lambda}{a}.$$

Choosing $f(x) = x^2$, we get

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda + \lambda^2}{a^2}.$$



Chebyshev's Inequality

This is Pafnuty's inequality:
Theorem:

$$Pr[|X - E[X]| > a] \leq \frac{var[X]}{a^2}, \text{ for all } a > 0.$$

Proof: Let $Y = |X - E[X]|$ and $f(y) = y^2$. Then,

$$Pr[Y \geq a] \leq \frac{E[f(Y)]}{f(a)} = \frac{var[X]}{a^2}.$$

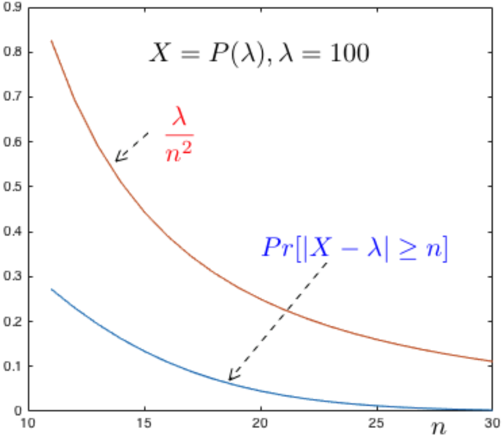
□

This result confirms that the variance measures the "deviations from the mean."

Chebyshev and Poisson

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $var[X] = \lambda$. Thus,

$$Pr[|X - \lambda| \geq n] \leq \frac{var[X]}{n^2} = \frac{\lambda}{n^2}.$$



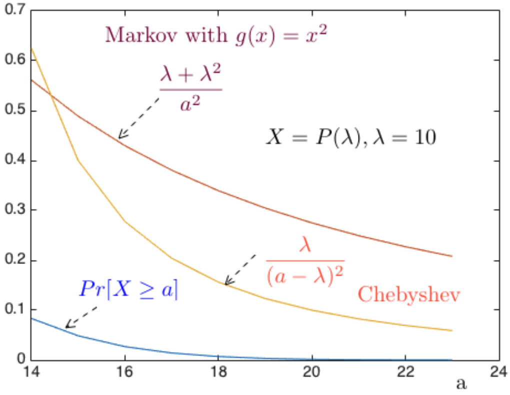
Chebyshev and Poisson (continued)

Let $X = P(\lambda)$. Then, $E[X] = \lambda$ and $var[X] = \lambda$. By Markov's inequality,

$$Pr[X \geq a] \leq \frac{E[X^2]}{a^2} = \frac{\lambda + \lambda^2}{a^2}.$$

Also, if $a > \lambda$, then $X \geq a \Rightarrow X - \lambda \geq a - \lambda > 0 \Rightarrow |X - \lambda| \geq a - \lambda$.

Hence, for $a > \lambda$, $Pr[X \geq a] \leq Pr[|X - \lambda| \geq a - \lambda] \leq \frac{\lambda}{(a-\lambda)^2}$.



Estimation: Expectation and Mean Squared Error.

"Best" guess about Y, is $E[Y]$.
If "best" is Mean Squared Error.

More precisely, the value of a that minimizes $E[(Y-a)^2]$ is $a = E[Y]$.

Proof:

Let $\hat{Y} := Y - E[Y]$.
Then, $E[\hat{Y}] = E[Y - E[Y]] = E[Y] - E[Y] = 0$.
So, $E[\hat{Y}c] = 0, \forall c$. Now,

$$\begin{aligned} E[(Y-a)^2] &= E[(Y - E[Y] + E[Y] - a)^2] \\ &= E[(\hat{Y} + c)^2] \text{ with } c = E[Y] - a \\ &= E[\hat{Y}^2 + 2\hat{Y}c + c^2] = E[\hat{Y}^2] + 2E[\hat{Y}c] + c^2 \\ &= E[\hat{Y}^2] + 0 + c^2 \geq E[\hat{Y}^2]. \end{aligned}$$

Hence, $E[(Y-a)^2] \geq E[(Y - E[Y])^2], \forall a$. $\square$

Estimation: Preamble

Thus, if we want to guess the value of Y, we choose $E[Y]$.

Now assume we make some observation X related to Y.

How do we use that observation to improve our guess about Y?

How? Conditional expectation.

Expectation: for random variable X for event A.

$Pr[X = x|A] = \frac{Pr[X=x \cap A]}{Pr[A]}$

Conditional Expectation: $E[X|A] = \sum_x x \times Pr[X = x|A]$.

Conditioned on event A, what prediction minimizes mean squared error (MMSE)? $E[X|A]$

For random variable X and Y.

$E[X|Y = y] = \sum_x x \times Pr[X = x|Y = y]$.

If you know y, what is MMSE prediction? $E[X|y]$.

Covariance is related to best linear predictor for X.

More on Tuesday.

Summary

Variance

- Variance: $var[X] := E[(X - E[X])^2] = E[X^2] - E[X]^2$
- Fact: $var[aX + b] a^2 var[X]$
- Sum: X, Y, Z pairwise ind. $\Rightarrow var[X + Y + Z] = \cdots$
- Markov (future): $Pr[X \geq a] \leq E[f(X)]/f(a)$ where ...
- Chebyshev (future): $Pr[|X - E[X]| \geq a] \leq var[X]/a^2$

Random Variables so far.

Probability Space: Ω, $Pr : \Omega \rightarrow [0,1]$, $\sum_{\omega \in \Omega} Pr(w) = 1$.
Random Variables: $X : \Omega \rightarrow R$.
Associated event: $Pr[X = a] = \sum_{\omega : X(\omega) = a} Pr(\omega)$
X and Y independent $\iff$ all associated events are independent.
Expectation: $E[X] = \sum_a a Pr[X = a] = \sum_{\omega \in \Omega} X(\omega) Pr(\omega)$.
Linearity: $E[X + Y] = E[X] + E[Y]$.

Variance: $Var(X) = E[(X - E[X])^2] = E[X^2] - (E(X))^2$
For independent X, Y, $Var(X + Y) = Var(X) + Var(Y)$.
Also: $Var(cX) = c^2 Var(X)$ and $Var(X + b) = Var(X)$.

Poisson: $X \sim P(\lambda)$ $E(X) = \lambda$, $Var(X) = \lambda$.
Binomial: $X \sim B(n,p)$ $E(X) = np$, $Var(X) = np(1-p)$
Uniform: $X \sim U\{1, \ldots, n\}$ $E[X] = \frac{n+1}{2}$, $Var(X) = \frac{n^2 - 1}{12}$.
Geometric: $X \sim G(p)$ $E(X) = \frac{1}{p}$, $Var(X) = \frac{1-p}{p^2}$