9/24: Expectation and Distributions

Announcements

Assignment 2 is due Thursday.

Assignment 3 will be out then.

Expectation

Intuitive notion: average value of a random variable.

$$E[X] = \sum_x x \, \mathsf{Pr}(X = x)$$

In an empirical sample, $x_1, x_2, x_3, \ldots, x_N$

$$\begin{aligned} \tilde{E}[X] &= \frac{1}{N} \sum_i x_i \\ &= \sum_x x \frac{\#(x_i = x)}{N} \end{aligned}$$

Note: Expectation of a probability distribution doesn't make sense

Continuous Case

$$E[X] = \int_{-\infty}^{\infty} x \, f_X(x) \, dx$$

Assuming the integral exists

Non-existant Expectation

Recall the distribution of gcds in random pairs of integers, approximated by

$$\begin{aligned} \mathsf{Pr}(k) &\propto \frac{1}{k^2} \\ &= \frac{6}{\pi^2 k^2} \end{aligned}$$

$$E[X] = \frac{6}{\pi^2} \sum_k \frac{1}{k}$$

Non-existant Expectation

Many "heavy tailed" distributions, which arise in the analysis of computing networks, do not have means.

"Power-law" (polynomial decay, Pareto-Lévy):

$$\mathsf{Pr}(X > x) = C\, x^{-\alpha}, \quad x > 0$$

For $0 < \alpha \leq 1$, mean (and variance) infinite

For $1 < \alpha < 2$, mean finite, variance infinite

General Case

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x\, dF_X(x) \\ &= \int_{-\infty}^{\infty} x\, d\,\mathsf{Pr}(X \leq x) \end{aligned}$$

Expectation is Linear

For random variables $X_1, X_2, \ldots, X_n$ and constants $a_1, a_2, \ldots a_n$

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

(assuming each expectation exists...)

Binomial

Suppose that $X \sim$ Binomial(n, p). Then $X = \sum_{i=1}^{n} B_i$ where $B_i \sim$ Bernoulli(p).

$$\begin{aligned} E(X) &= E(\sum_i B_i) \\ &= \sum_i p \\ &= n\,p \end{aligned}$$

A Standard Trick

Can derive it "directly" using differentiation

$$\begin{aligned} E(X) &= \sum_k k \binom{n}{k} p^k q^{n-k} \\ &= p \frac{\partial}{\partial p} \sum_k \binom{n}{k} p^k q^{n-k} \\ &= p \frac{\partial}{\partial p} (p+q)^n \\ &= np \end{aligned}$$

(It's no accident that this is $\frac{\partial}{\partial p} \log$ (normalizing term))

Example

n letters, each with a matching addressed envelope. Letters are placed randomly into envelopes. What is expected number of letters that end up in the correct envelope?

Example

Let $X_i = 1$ if the i-th letter ends up in it's correct envelope, $X_i = 0$ otherwise. Then Pr$(X_i = 1) = \frac{1}{n}$.

Bernoulli, with identical distribution. Independent?

Example

$$E\left(X_1 + X_2 + \cdots X_n\right) = \frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} = 1$$

(independent of n)

Products

If $X_1, X_2, \ldots, X_n$ are *independent* then

$$E\left(\prod_i X_i\right) = \prod_i E\left(X_i\right)$$

Products

If $X_1, X_2, \ldots, X_n$ are *independent* then

$$\begin{aligned}
E\left(\prod_i X_i\right) &= \sum_{v_1, v_2, \ldots, v_n} v_1 v_2 \cdots v_n \, \mathsf{Pr}\left(\cap_i \{X_i = v_i\}\right) \\
&= \sum_{v_1, v_2, \ldots, v_n} v_1 v_2 \cdots v_n \prod_i \mathsf{Pr}(\{X_i = v_i\}) \\
&= \prod_i \sum_{v_i} v_i \, \mathsf{Pr}(\{X_i = v_i\})
\end{aligned}$$

Expectations of Functions of a R.V.

If $Y = f(X)$ then

$$E(Y) = \int_{-\infty}^{\infty} r(x)\, dF_X(x) dx$$

Example: Take a stick of unit length and divide it into two pieces. What is the average length of the longer piece?

Expectations of Functions of a R.V.

$$\begin{aligned} E(Y) &= \int_0^{1/2} (1-x)\,dx + \int_{1/2}^1 x\,dx \\ &= \frac{3}{4} \end{aligned}$$

Example

Suppose that X and Y are independent with exponential distributions with parameter 1. What is

$$E\left[\exp\left(\frac{1}{2}(X+Y)\right)\right]?$$

Example

$$\begin{aligned} E\left[\exp\left(\frac{1}{2}(X+Y)\right)\right] &= E\left[\exp\left(\frac{1}{2}X\right)\exp\left(\frac{1}{2}Y\right)\right] \\ &= \left(E\left[\exp\left(\frac{1}{2}X\right)\right]\right)^2 \\ &= \left(\int_0^\infty e^{\frac{1}{2}x} e^{-x}\,dx\right)^2 \\ &= 4 \end{aligned}$$

Example: Geometric

What is the expected number of flips need to obtain the first head with coin of bias p?

Example: Geometric

What is the expected number of flips need to obtain the first head with coin of bias p?

$$\begin{aligned} E(X) &= p + 2(1-p)p + 3(1-p)^2 p + \cdots \\ &= p \sum_{i=1}^{\infty} i q^{i-1} \\ &= p \frac{d}{dq}\left(\frac{1}{1-q}\right) \\ &= p \frac{1}{p^2} \\ &= \frac{1}{p} \end{aligned}$$

Example: Geometric

If a r.v. takes on integer values $n \geq 0$ then

$$E(X) = \sum_k \mathrm{Pr}(X \geq k)$$

So, for the geometric distribution

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} \mathrm{Pr}(X \geq k) \\ &= \sum_{k=0}^{\infty} q^k \\ &= \frac{1}{1-q} \\ &= 1/p \end{aligned}$$

Variance

The *variance* of a random variable X is the expectation of $(X - E(X))^2$:

$$\begin{aligned} \mathrm{Var}(X) &= E\left[(X - E[X])^2\right] \\ &= E\left[X^2 - E[X]^2\right] \end{aligned}$$

whenever this expectation is well-defined.

Measure of spread or dispersion around the mean.

Variance

The *standard deviation* is the square root of the variance:

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

(Same "units" as X)

Variance – Properties

If $X_1, X_2, \ldots, X_n$ are independent, then

$$\begin{aligned}\mathrm{Var}(X_1 + X_2 + \cdots + X_n) &= \\ &\mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n)\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(a_1X_1 + a_2X_2 + \cdots + a_nX_n) &= \\ &a_1^2\mathrm{Var}(X_1) + a_2^2\mathrm{Var}(X_2) + \cdots + a_n^2\mathrm{Var}(X_n)\end{aligned}$$

Example: Binomial

$X \sim \mathrm{Bernoulli}(p)$ has mean p and variance

$$\begin{aligned}\mathrm{Var}(X) &= E(X^2) - E(X)^2 \\ &= p - p^2 \\ &= p(1-p) \\ &= pq\end{aligned}$$

If $X \sim \mathrm{Binom}(n, p)$ then

$$\mathrm{Var}(X) = np(1-p)$$

Covariance

For jointly distributed r.v.s X and Y the *covariance* is given by

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E\left[(X - E(X))(Y - E(Y))\right] \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

By the Schwarz inequality

$$E(XY) \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}$$

and thus $|\mathrm{Cov}(X,Y)| \leq \sigma_X\sigma_Y$. So, the *correlation*

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}$$

is between -1 and 1.

Covariance

- A measure of how X and Y vary together.
- If X is large when Y is, or X is small when Y is large, then $|\mathrm{Cov}(X,Y)|$ will be large
- Correlation $\rho(X,Y)$ scales to $[-1, 1]$.

Sample Statistics

The *sample mean* of $X_1, X_2, \ldots, X_n$ is

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

The *sample variance* is defined as

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

This is the "right" definition since if X_i are independent and distributed as X then

$$E(\overline{X}_n) = E(X) \text{ and } E(S_n^2) = \mathrm{Var}(X)$$

Sample Statistics

Note that if X_i are iid then $\mathrm{Var}(\overline{X}_n) = \frac{1}{n}\sigma_X^2$

The probability of deviations from the mean can be estimated using some simple inequalities.

Markov inequality

If $\Pr(X \geq 0) = 1$ then $\Pr(X \geq t) \leq \frac{E(X)}{t}$.

Proof:
$$\begin{aligned} E(X) &= \sum_x x\, f(x) \\ &= \sum_{x<t} x\, f(x) + \sum_{x \geq t} x\, f(x) \\ &\geq \sum_{x \geq t} x\, f(x) \\ &\geq t \sum_{x \geq t} f(x) \\ &= t \Pr(X \geq t) \end{aligned}$$

Same proof for continuous case.

Chebyshev Inequality

If $\mathrm{Var}(X)$ exists then $\Pr(|X - E(X)| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$.

Proof: By the Markov inequality

$$\begin{aligned} \Pr(|X - E(X)| \geq t) &= \Pr((X - E(X))^2 \geq t^2) \\ &\leq \frac{E((X - E(X))^2)}{t^2} \\ &= \frac{\mathrm{Var}(X)}{t^2} \end{aligned}$$

Bounds on Sample Mean

Intuitively, for iid random variables the sum gets more concentrated on the mean. Chebyshev quantifies this as

$$\begin{aligned} \Pr(|\overline{X}_n - E(X)| \geq t) &\leq \frac{\mathrm{Var}(\overline{X}_n)}{t^2} \\ &= \frac{\sigma^2}{nt^2} \end{aligned}$$

Bounds on Sample Mean

If the standard deviation is 2, how large should n be so that $\Pr(|\overline{X}_n - E(X)| < 1)$ is at least 0.99?

Bounds on Sample Mean

How large should n be so that $\Pr(|\overline{X}_n - E(X)| < 1)$ is at least 0.99?

By Chebyshev,

$$\Pr(|\overline{X}_n - E(X)| \geq 1) \leq \frac{2^2}{n} \leq 0.01$$

if $n \geq 400$.

Bounds on Sample Mean

How many times do we need to flip a fair coin to ensure that the probability is at least 0.7 that the proportion of heads lies in the interval $[0.4, 0.6]$?

Bounds on Sample Mean

For n flips, the mean and variance are $n/2$ and $n/4$. Thus, by Chebyshev, if H is the number of heads,

$$\begin{aligned} \Pr(0.4 \le \overline{X}_n \le 0.6) &= \Pr(0.4n \le H \le 0.6n) \\ &= \Pr\left(|H - \frac{n}{2}| \le 0.1n\right) \\ &\ge 1 - \frac{n}{4\,(0.1n)^2} \\ &= 1 - \frac{25}{n} \end{aligned}$$

From this, we need to choose $n \ge 84$.

Bounds on Sample Mean

However, from directly computing quantiles of the binomial, we find that 15 tosses is enough!

Compromise of a general (non-parametric) test.

Better bounds are commonly used in machine learning and statistics: Hoeffding's inequality (Gaussian approximation), relative Chernoff bounds.

Conditional Distributions

For discrete random variables X and Y define the *conditional probability mass function* as

$$\begin{aligned} f_{X|Y}(x\,|\,y) &= \Pr(X = x\,|\,Y = y) \\ &= \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)} \\ &= \frac{f_{X,Y}(x, y)}{f_Y(y)} \end{aligned}$$

whenever $f_Y(y) > 0$.

In the continuous case, we make the same definition "formally"

Conditional Expectation

For discrete random variables, the *conditional expectation* of Y given $X = x$ is

$$E(Y\,|\,X = x) = \sum_y y \Pr(Y = y\,|\,X = x)$$

In the continuous case,

$$E(Y\,|\,X = x) = \int y \Pr(Y = y\,|\,X = x)$$

Semantics: $E(Y\,|\,X)$ is a random variable.

Example

Suppose that $X \sim$ Uniform(0,1) and given $X = x$, $Y \mid X \sim$ Uniform$(x, 1)$. Then

$$
\begin{aligned}
E(Y \mid X = x) &= \int_x^1 y \, f_{Y \mid X}(y \mid x) \, dy \\
&= \int_x^1 y \, \frac{1}{1-x} \, dy \\
&= \frac{1}{2} \cdot \frac{1-x^2}{1-x} \\
&= \frac{1+x}{2}
\end{aligned}
$$

Iterated Expectations

We can always condition on a random variable and then marginalize it away:

Theorem. For random variables X and Y, and any function $f(\cdot,\cdot)$ we have

$$E(Y) = E\left[E(Y \mid X)\right]$$

$$E(r(X,Y)) = E\left[E(r(X,Y) \mid X)\right]$$

(An important and commonly used conceptual and computational technique...)

Iterated Expectations

Proof (in discrete case): Remembering $E(Y \mid X)$ is a r.v.:

$$
\begin{aligned}
E(E(Y \mid X)) &= \sum_x \mathrm{Pr}(X = x) \sum_y y \, \mathrm{Pr}(Y = y \mid X = x) \\
&= \sum_x \mathrm{Pr}(X = x) \sum_y y \, \frac{\mathrm{Pr}(Y = y, X = x)}{\mathrm{Pr}(X = x)} \\
&= \sum_x \sum_y y \, \mathrm{Pr}(Y = y, X = x) \\
&= \sum_y y \sum_x \mathrm{Pr}(Y = y, X = x) \;=\; E(Y)
\end{aligned}
$$

Note that $E(Y \mid X = x)$ is the random variable we're taking expectations of in the first line.

Iterated Expectations

Proof (in discrete case):

$$
\begin{aligned}
E(r(X,Y)) &= \sum_{x,y} f_{X,Y}(x,y) \, r(x,y) \\
&= \sum_x f_X(x) \sum_y f_{Y \mid X}(y \mid x) \, r(X,Y) \\
&= E\left[E(r(X,Y) \mid X)\right]
\end{aligned}
$$