QBS 120- Lecture 4
Expected values and variance
(Rice Chapter 4.1-4.2)

Rob Frost

Outline

- Expected value of a RV
- Expected value of functions of RVs
- Expectation of linear combinations of RVs
- Variance and standard deviation

Expected value of discrete RVs

For discrete RV X with PMF $p(x)$, the expectation of X (E[X], mean or μ_x) is defined as:

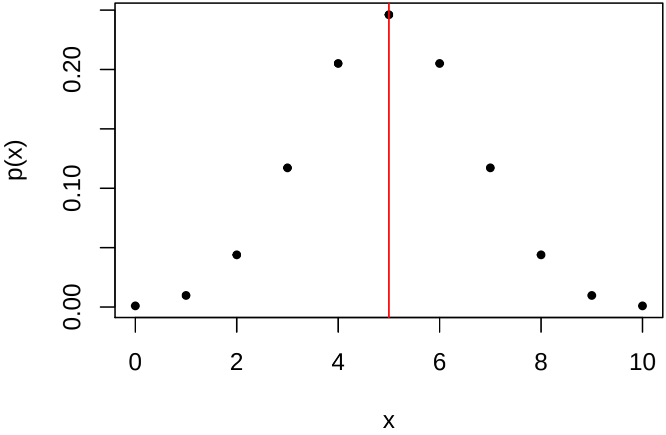
$$E[X] = \sum_{x_i \in X(\Omega)} x_i p(x_i)$$

assuming $\sum_{x_i \in X(\Omega)} |x_i| p(x_i) < \infty$.

Can think of E[X] as the center of mass of the frequency function X.

Expected value illustration

PMF and expected value for Binomial with n=10, p=0.5:



Expected value of Bernoulli RVs

If X is a Bernoulli RV with probability p, what is $E[X]$?

$$\begin{aligned} E[X] &= \sum_{x_i \in X(\Omega)} x_i p(x_i) \\ &= 1(p) + 0(1-p) \\ &= p \end{aligned}$$

Expected value of Poisson RVs

If X is a Poisson RV with parameter λ, what is $E[X]$?

$$\begin{aligned}
E[X] &= \sum_{x_i \in X(\Omega)} x_i p(x_i) \\
&= \sum_{k=1}^{\infty} k\left(\frac{\lambda^k e^{-\lambda}}{k!}\right) = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda}(e^{\lambda}) \\
&= \lambda
\end{aligned}$$

This makes sense when we remember that Poisson RVs approximate binomial RVs for large n and p with $\lambda = np$. The expected value of binomial RVs is the intuitively expected np.

Expected value of continuous RVs

For continuous RV X with density $f(x)$, the expectation of X is defined as:

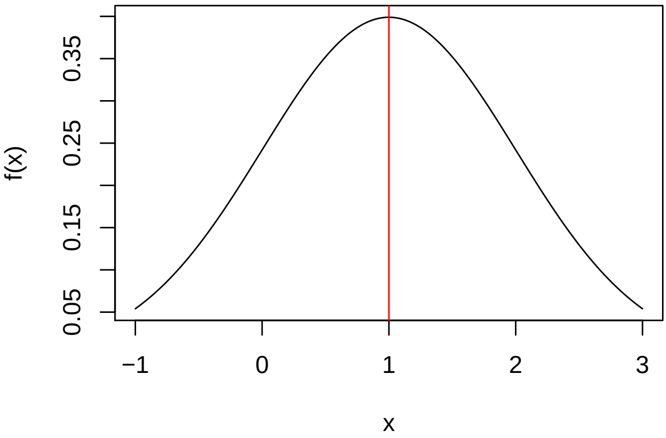
$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

assuming $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$.

Can again think of E[X] as the center of mass of the density of X.

Expected value illustration

Density and expected value for normal with $\mu = 1$, $\sigma^2 = 1$



Expected value of normal RV

If X is a normal RV with parameters μ and σ, what is $E[X]$?

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-0.5(x-\mu)^2/\sigma^2} dx
\end{aligned}
$$

Make a change of variables $z = x - \mu$:

$$
\begin{aligned}
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (z+\mu) e^{-0.5(z)^2/\sigma^2} \frac{d}{dx}(x-\mu) dz \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-0.5(z)^2/\sigma^2} dz + \frac{\mu}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-0.5(z)^2/\sigma^2} dz \\
&= 0 + \mu = \mu
\end{aligned}
$$

Expected value of a Cauchy RV

Cauchy RVs can be formed as the ratio of independent standard normal RVs. The Cauchy density is:

$$f(x) = \frac{1}{\pi}\left(\frac{1}{1+x^2}\right)$$

Since this density is symmetric about 0, would think $E[X] = 0$, however:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x}{1+x^2} dx \end{aligned}$$

However, $\int_{-\infty}^{\infty} \frac{|x|}{1+x^2} dx = \infty$ so the expectation does not exist.

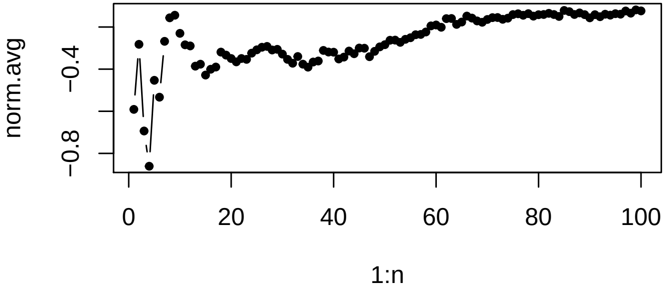
Average of independent RVs

As we'll learn in Chapter 5, the average of i.i.d. (independent and identically distributed) RVs converges to the expected value as $n \to \infty$. Let's simulate that for both Cauchy and standard normal RVs.

```
> set.seed(11)
> n = 100
> norm.avg = cumsum(rnorm(n))/(1:n)
> cauchy.avg = cumsum(rnorm(n)/rnorm(n))/(1:n)
```

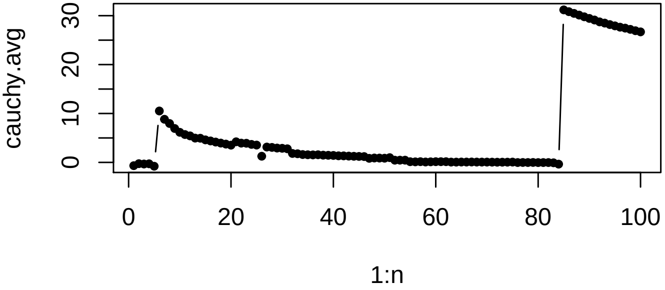
Average of independent standard normal

```
> plot(1:n, norm.avg, type="both", pch=20)
```



Average of independent Cauchy

```
> plot(1:n, cauchy.avg, type="both", pch=20)
```



Markov Inequality

If X is an RV with $P(X \geq 0) = 1$ and $E[X]$ exists,
then $P(X \geq t) \leq E[X]/t$ (i.e., prob. that X is larger than E[X] is small)

Proof (for discrete case):

$$E[X] = \sum_{x \in X(\Omega)} xp(x)$$

$$E[X] = \sum_{x<t} xp(x) + \sum_{x \geq t} xp(x)$$

Note: all terms positive

$$E[X] \geq \sum_{x \geq t} xp(x) \geq \sum_{x \geq t} tp(x) = tP(X \geq t)$$

$$\frac{E[X]}{t} \geq P(X \geq t) \quad \square$$

Markov Inequality, example

Check for Poisson RV with $\lambda = 1$.

- Know that $E[X] = \lambda = 1$
- What is $P(X \geq 2)$?
- From Markov Inequality, know that $P(X \geq 2) \leq E[X]/t = 0.5$
- From the Poisson CDF, know that $P(X \geq 2) = 1 - F(2) = 0.080301397071 3942$

Conclusion: Markov Inequality holds for all positive RVs with a valid expectation but is a very conservative bound.

Outline

- Expected value of a RV
- **Expected value of functions of RVs**
- Expectation of linear combinations of RVs
- Variance and standard deviation

Theorem A, Rice 4.1.1

Let $Y = g(X)$ for RV X. What is $E[Y]$?

- X is discrete with pmf $p(x)$:

$$E[Y] = \sum_{x \in X(\Omega)} g(x)p(x)$$

  if $\sum |g(x)|p(x) < \infty$

- X is continuous with density $f(x)$:

$$E[Y] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

  if $\int |g(x)|f(x)dx < \infty$

Theorem B, Rice 4.1.1

Theorem A naturally extends to a function of jointly distributed RVs: $Y = g(X_1, ..., X_n)$

- X_i are discrete:

$$E[Y] = \sum_{x_1, ..., x_n} g(x_1, ..., x_n) p(x_1, ..., x_n)$$

  if $\sum |g(x_1, ..., x_n)| p(x_1, ..., x_n) < \infty$

- X_i are continuous:

$$E[Y] = \int ... \int g(x_1, ..., x_n) f(x_1, ..., x_n) dx_1 ... dx_n$$

  if $\int ... \int |g(x_1, ..., x_n)| f(x_1, ..., x_n) dx_1 ... dx_n < \infty$

Example B, Rice 4.1.1

A stick 1 foot long is broken randomly in two places. What is the average length of the middle piece?

- Treat the two break points as independent standard uniform RVs: $X_1, X_2 \sim U(0,1)$
- Let Y be a RV representing the length of the middle piece. Y can be defined as a function of X_1 and X_2: $Y = g(X_1, X_2) = |X_1 - X_2|$
- According to Theorem B, $E[Y]$ is:

$$
\begin{aligned}
E[Y] &= \int_0^1 \int_0^1 |x_1 - x_2| f_{X_1,X_2}(x_1, x_2) dx_1 dx_2 \\
&= \int_0^1 \int_0^1 |x_1 - x_2| dx_1 dx_2 \\
&= \int_0^1 \int_0^{x_1} (x_1 - x_2) dx_2 dx_1 + \int_0^1 \int_{x_1}^1 (x_2 - x_1) dx_2 dx_1 = 1/3
\end{aligned}
$$

Corollary A, Rice 4.1.1

If X and Y are independent RVs and f() and g() are functions, then:

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$$

This covers the simple case of the expectation of a product of independent RVs:

$$E[XY] = E[X]E[Y]$$

Outline

- Expected value of a RV
- Expected value of functions of RVs
- **Expectation of linear combinations of RVs**
- Variance and standard deviation

Theorem A, Rice 4.1.2

Define RV Y as a linear function of jointly distributed RVs $X_1, ..., X_n$:

$$Y = a + \sum_{i=1}^{n} b_i X_i$$

In this case, $E[Y]$ has a particularly nice form:

$$E[Y] = a + \sum_{i=1}^{n} b_i E[X_i]$$

$\rightarrow$ The expectation of a sum of RVs is the sum of the expectations. Does not require that the RVs are independent!

Expected value of binomial RVs

If Y is a binomial RV with parameters n and p, what is $E[Y]$?

$$
\begin{aligned}
E[Y] &= \sum_{x \in X(\Omega)} xp(x) \\
&= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}
\end{aligned}
$$

How to evaluate? Trick: realize that a binomial is a sum of independent Bernoulli RVs. So, $E[Y]$ is sum of the expectations of n Bernoulli RVs $X_i, \dots, X_n$ all with parameter p:

$$
\begin{aligned}
E[Y] &= \sum_{i=1}^{n} E[X_i] \\
&= \sum_{i=1}^{n} p \\
&= np
\end{aligned}
$$

DNA motif counting (Example D, Rice 4.1.2)

Given a segment of sequenced DNA, how to identify non-random motifs, i.e., short sequences that occur more frequently than would be expected at random? Approach:

- Assume that the DNA sequence is random: the bases A, C, G, T occur with equal likelihood at each location and each location is independent. Let B_n represent the base at position n.
- Given a specific motif (e.g., TATA), we count the number of occurrences in a given sequence including overlaps (e.g., TATA occurs 3 times in ACTATATAGATATA).
- For a sequence of length N and motif of length q, let the indicator variable I_n represent whether the motif starts at position n. Let M_k represent the base at position k in the motif. Under the random sequence assumption we can use the multiplication law to find $P(I_n = 1)$:

$$\begin{aligned} P(I_n = 1) &= P(B_n = M_1 \cap ... \cap B_{n+q-1} = M_q) \\ &= P(B_n = M_1)...P(B_{n+q-1} = M_q) = (\frac{1}{4})^q \end{aligned}$$

DNA motif counting, continued

- The total number of occurrences of the motif (W) is the sum of the I_n values for all potential starting locations ($i = 1, ..., N - q + 1$):

$$W = \sum_{i=1}^{N-q+1} I_n$$

- What is the expected number of occurrences, E[W]? Can use the linearity of expectations to compute:

$$\begin{aligned} E[W] &= \sum_{i=1}^{N-q+1} E[I_n] \\ &= \sum_{i=1}^{N-q+1} 1(\frac{1}{4}^q) + 0(1 - \frac{1}{4}^q) \\ &= (N - q + 1)(\frac{1}{4})^q \end{aligned}$$

Note: this works even though the I_n are not independent!

DNA motif counting, continued

- What is E[W] for motif TATA and sequence ACTATATAGATATA?
- In this case, N=14 and q=4. So,

$$\begin{aligned} E[W] &= (N-q+1)(\frac{1}{4})^q \\ &= (14-4+1)(0.25)^4 \\ &= 0.043 \end{aligned}$$

- The actual count is 3 which is much larger than the expected count under the random sequence assumption. This certainly suggests that the motif TATA is appearing in a non-random fashion in this sequence but how to quantify this unexpected outcome?

This type of question is addressed via hypothesis testing and quantified using p-values. We'll cover all of that in Chapter 9.

Outline

- Expected value of a RV
- Expected value of functions of RVs
- Expectation of linear combinations of RVs
- **Variance and standard deviation**

Variance

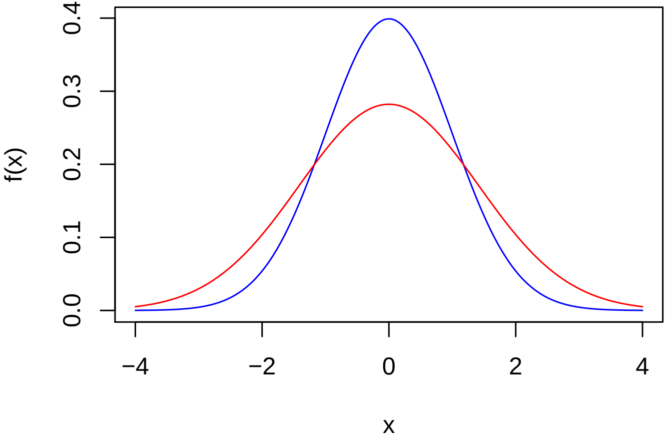
For RV X with expectation E[X], the variance of X is:

$$Var(X) = \sigma_X^2 = E[(X - E[X])^2]$$

In other words, the variance is the expected value of the square of the distance between X and E[X]. It is a measure of the dispersion of the density of X around E[X].

Variance illustration

Density normal RVs with $\mu = 0$ and either $\sigma^2 = 1$ (blue) or $\sigma^2 = 2$ (red)



Variance, continued

For discrete RV X with PMF p(x) and expectation μ, the variance is:

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] \\ &= E[(X - \mu)^2] \\ &= \sum_{x \in X(\Omega)} (x - \mu)^2 p(x) \end{aligned}$$

For continuous RV X with density f(x), the variance is:

$$Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Standard deviation

Standard deviation is the positive square root of the variance:

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

If the RV X is measured in units of a (e.g., kg):

- Variance has units of a^2
- Standard deviation has units of a.

Because σ_X has the same units as X, it can be easier to interpret. Useful rule of thumb: For normally distributed data, approximately $2/3$ of the data are within 1 sd of the mean.

Variance of Bernoulli RVs

X is a Bernoulli RV with parameter p. The variance of X is:

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] \\ &= E[(X - p)^2] \\ &= E[X^2 - 2pX + p^2] \end{aligned}$$

Treat $g(X) = X^2 - 2pX + p^2$ and find E[g(X)]:

$$\begin{aligned} &= (1 - 2p + p^2)p + (p^2)(1 - p) \\ &= p - 2p^2 + p^3 + p^2 - p^3 \\ &= p - p^2 \\ &= p(1 - p) \end{aligned}$$

Variance is invariant to changes in location

If X is an RV and Var(X) exists, the addition of a constant does not change the variance:

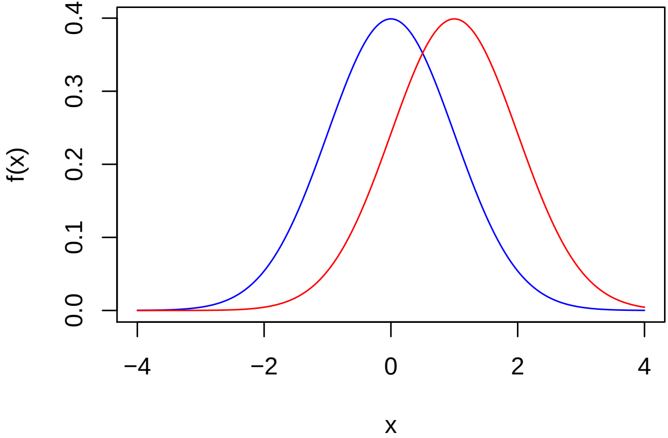
$$Var(a + X) = Var(X)$$

Proof:

$$\begin{aligned} Var(a + X) &= E[(a + X - E[a + X])^2] \\ &= E[(a + X - a - E[X])^2] \\ &= E[(X - E[X])^2] \\ &= Var(X) \end{aligned}$$

Location change illustration

Density normal RVs with $\sigma^2 = 1$ and either $\mu = 0$ (blue) or $\mu = 1$ (red):



Variance of a linear transformation

If X is an RV and Var(X) exists, the variance of linear transformation $Y = a + bX$ is:

$$Var(Y) = b^2 Var(X)$$

Proof:

$$\begin{aligned}
Var(Y) &= E[(Y - E[Y])^2] \\
&= E[(a + bX - a - bE[X])^2] \\
&= E[b^2(X - E[X])^2] \\
&= b^2 E[(X - E[X])^2] \\
&= b^2 Var(X)
\end{aligned}$$

Alternate formula for variance (Theorem B, Rice 4.2)

Variance can also be defined as:

$$Var(X) = E[X^2] - E[X]^2$$

Proof (let $E[X] = \mu$):

$$\begin{aligned}
Var(X) &= E[(X-\mu)^2] \\
&= E[X^2 - 2\mu X + \mu^2] \\
&= E[X^2] - E[2\mu X] + E[\mu^2] \\
&= E[X^2] - 2\mu E[X] + \mu^2 \\
&= E[X^2] - 2\mu^2 + \mu^2 \\
&= E[X^2] - \mu^2
\end{aligned}$$

Variance of uniform RV

Let X be a standard uniform RV. What is Var(X)?

$$
\begin{aligned}
Var(X) &= E[X^2] - E[X]^2 \\
&= E[X^2] - (1/2)^2 \\
&= \int_0^1 g(x)f(x)dx - 1/4 \\
&= \int_0^1 x^2 dx - 1/4 \\
&= 1/3 - 1/4 \\
&= 1/12
\end{aligned}
$$

Chebyshev's inequality

Let X be a RV with $E[X] = \mu$ and $Var(X) = \sigma^2$. For any $t > 0$:

$$P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$$

Proof:

- Let $Y = (X - \mu)^2$
- $E[Y] = E[(X - \mu)^2] = Var(X) = \sigma^2$
- Apply Markov's inequality to Y and $z = t^2 > 0$:

$$\begin{aligned} P(Y \geq z) &\leq E[Y]/z \\ P(Y > z) &\leq E[Y]/z \\ P((X - \mu)^2 > z) &\leq \sigma^2/z \\ P(|X - \mu| > \sqrt{z}) &\leq \sigma^2/z \\ P(|X - \mu| > t) &\leq \sigma^2/t^2 \quad \square \end{aligned}$$

Corollary A

If Var(X) $= 0$, then $P(X = \mu) = 1$
(i.e., if X has 0 variance, it is a constant)

Proof by contradiction:

- Suppose $P(X = \mu) < 1$.
- Then for some $\varepsilon > 0$, $P(|X - \mu| \geq \varepsilon) > 0$ (i.e., there must be some density away from μ)
- However, according to Chebyshev's inequality:

$$P(|X - \mu| \geq \varepsilon) = Var(X)/\varepsilon^2 = 0$$

Measurement error

The error in scientific instruments can be decomposed into both systematic (i.e., repeatable) and random error. If the value output by the instrument is represented by the RV X, measurement error can be modeled using the following equation:

$$X = x_0 + \beta + \varepsilon$$

where:

- X is the value reported by the instrument
- x_0 is the true value (a constant)
- β is the systematic error or bias of the instrument (a constant)
- ε is the random error is a represented as a RV with $E[\varepsilon] = 0$ and $Var(\varepsilon) = \sigma^2$

Note: linear regression models have a very similar form, more on that in QBS 121.

Measurement error, continued

According to the measurement error model

- The average value returned by the instrument is only impacted by the bias:

$$\begin{aligned} E[X] &= E[x_0 + \beta + \varepsilon] \\ &= E[x_0] + E[\beta] + E[\varepsilon] \\ &= x_0 + \beta \end{aligned}$$

- The variation in the measured value is all due to the random error:

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] \\ &= E[(x_0 + \beta + \varepsilon - x_0 - \beta)^2] \\ &= E[\varepsilon^2] \\ &= Var(\varepsilon) \\ &= \sigma^2 \end{aligned}$$

Measurement error, continued

A common representation of measurement error is the mean squared error (MSE):

$$MSE = E[(X - x_0)^2]$$

Which can be decomposed into the error variance plus the squared bias:

$$MSE = \beta^2 + \sigma^2$$

Proof:

$$\begin{aligned}
Var(X - x_0) &= E[(X - x_0)^2] - E[X - x_0]^2 \\
Var(X - x_0) &= MSE - E[X - x_0]^2 \\
MSE &= Var(X - x_0) + E[X - x_0]^2 \\
&= Var(X - x_0) + E[X - x_0]^2 \\
&= Var(X) + (E[X] - E[x_0])^2 \\
&= \sigma^2 + (x_0 + \beta - x_0)^2 \\
&= \sigma^2 + \beta^2
\end{aligned}$$