Math E-156 36 Proofs

My rankings of each proof's length from the **revised** list of 36 using Paul's Ballot box (which is now iSites!)

Group 1		Vote (3+, 3 – in each group)	Comments
1	1-1		Medium & straightforward
2	2-1		Short & straightforward
3	2-2		Long & complex
4	3-1		Short & straightforward
5	3-2		Short
6	4-1		Medium
7	4-2		Long & complex
8	4-3		Medium
9	5-1		Long & complex
Group 2			
10	5-2		Short
11	5-3		Short & straightforward
12	6-1		Medium & straightforward
13	6-2		Short & straightforward
14	7-1		Long & complex
15	7-2		Medium & complex
16	7-3		Medium
17	8-1		Long
18	8-2		Long & complex
Group 3			
19	8-3		Long & complex
20	8-4		Medium
21	8-5		Medium but check requested
22	8-6		Long & complex
23	8-7		Short but check requested, straightforward
24	8-8		Medium & straightforward
25	9-1		Medium & straightforward
26	9-2		Medium
27	9-3		Long & complex
Group 4			
28	9-4		Long and check requested, complex
29	9-5		Medium & complex
30	10-1		Short & straightforward
31	10-2		Short & straightforward
32	10-3		Medium
33	11-1		Medium
34	11-2		Long
35	12-1		Short & straightforward
36	12-2		Long & complex

1. 1-1

The variance $\mathrm{Var}[X]$ of a random variable X is defined as $E[(X - E[X])^2]$. Given that $E[a_1X_1 + a_2X_2] = a_1E[X_1] + a_2E[X_2]$ in all cases and that $E[X_1X_2] = E[X_1]E[X_2]$ for independent random variables, prove that

   (a) $\mathrm{Var}[X] = E[X^2] - (E[X])^2$.
   (b) $\mathrm{Var}[aX + b] = a^2\ \mathrm{Var}[X]$.
   (c) If X_1 and X_2 are independent, $\mathrm{Var}[X_1 + X_2] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2]$

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

$= E[X^2 - 2E[X]X + E[X]^2]$

Using $E[E[X]X] = E[X]^2$

$= E[X^2] - 2\,E[X]^2 + E[X]^2$

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

■

$Var[aX + b] = E[(aX + b)^2] - E[aX + b]^2$

$= E[a^2X^2 + 2abX + b^2] - (aE[X] + b)^2$

$= a^2E[X^2] + 2abE[X] + b^2 - a^2\,E[X]^2 - 2abE[X] - b^2$

$= a^2E[X^2] - a^2\,E[X]^2 = a^2Var[X]$

■

$Var[X_1 + X_2] = E[(X_1 + X_2)^2] - (E[X_1] + E[X_2])^2$

$= E[X_1^2] + E[2X_1X_2] + E[X_2^2] - E[X_1]^2 - 2E[X_1]E[X_2] - E[X_2]^2$

$E[2X_1X_2] = 2E[X_1]E[X_2]$ only when independent, so they cancel here

$= \ E[X_1^2] - E[X_1]^2 + E[X_2^2] - E[X_2]^2 = Var[X_1] + Var[X_2]$

■

2. 2-1 Variance of the mean of n independent random variables

This is a theorem of probability.

Suppose that $X_1, X_2, \cdots X_n$ are independent random variables, all with the same expectation μ and variance σ^2. Their mean is

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots X_n).$$

Prove that

$$E[\overline{X}] = \mu$$

and that

$$\mathrm{Var}[\overline{X}] = \frac{1}{n}\sigma^2.$$

$$E[\bar{X}] = \frac{E[X_1 + X_2 + \cdots X_n]}{n} = \frac{E[X_1] + E[X_2] + \cdots E[X_N]}{n} = \frac{n\mu}{n} = \mu$$

■

$$Var[X_1 + X_2 + \cdots X_n] = Var[X_1] + Var[X_1] + \cdots \; Var[X_N] = n\sigma^2$$

This needed the independence assumption.

$$Var[\bar{X}] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

■

3. 2-2

This is a theorem of statistics.

Let $X_1, X_2, \cdots X_n$ be independent random variables from a distribution with $\text{Var}[X_i] = \sigma^2 < \infty$.

We do not know the expectation μ, although we know from the previous result that the expectation of $\overline{X}$ is equal to μ.

We also do not know the variance. We try to estimate it by using the usual formula but, not knowing μ, we can do no better than to use $\overline{X}$ in its place.

$$\text{Prove that } E[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2] = \frac{n-1}{n}\sigma^2.$$

It follows that $S^2 = E[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2] = \sigma^2$. This is what var() computes.

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]$$

$$= E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2\bar{X}X_i + \bar{X}^2\right)\right]$$

$$= E\left[\frac{1}{n}\left(\sum_{i=1}^{n}X_i^2 - 2\bar{X}\sum_{i=1}^{n}X_i + \bar{X}^2\sum_{i=1}^{n}1\right)\right]$$

$$= E\left[\frac{1}{n}\left(\sum_{i=1}^{n}X_i^2 - 2\bar{X}(n\bar{X}) + \bar{X}^2 n\right)\right]$$

$$= \frac{1}{n}E\left[\sum_{i=1}^{n}X_i^2 - \bar{X}^2 n\right]$$

$$= \frac{1}{n}E\left(\left[\sum_{i=1}^{n}X_i^2\right] - nE[\bar{X}^2]\right)$$

Where we can break up the expectation because of linearity

$Var[X_i] = E[X_i^2] - E[X_i]^2$ or $\sigma^2 = E[X_i^2] - \mu^2$, so $E[X_i^2] = \sigma^2 + \mu^2$

$$E\left(\left[\sum_{i=1}^{n}X_i^2\right]\right) = n(\sigma^2 + \mu^2)$$

$Var[\bar{X}] = E[\bar{X}^2] - E[\bar{X}]^2$ or $\frac{\sigma^2}{n} = E[\bar{X}^2] - \mu^2$, so $E[\bar{X}^2] = \frac{\sigma^2}{n} + \mu^2$

Combining those properties into the expectation equation

$$\frac{1}{n}E\left(\left[\sum_{i=1}^{n} X_i^2\right] - nE[\bar{X}^2]\right) = \frac{1}{n}\left[n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right] = \frac{(n-1)\sigma^2}{n}$$

Similarly,

$$s^2 = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \sigma^2$$

■

4. 3-1

Prove that the sum of n independent Bernoulli random variables, each with parameter p, is a binomial random variable $Y \sim \text{Binom}(n, p)$, and that

$$E[Y] = np, \ \text{Var}\ Y = np(1-p).$$

(a) A Bernoulli random variable X has the value 1 with probability p, 0 with probability $1-p$. Calculate its expectation and variance.

(b) (The easy way) A binomial random variable Y is the sum of n independent Bernoulli random variables: $Y = X_1 + X_2 + \cdots + X_n$. Calculate its expectation and variance from this property alone.

(c) If Y has the value r, then r of the X_i have the value 1, $n-r$ have the value 0. Calculate the probability of this event, which can happen in many ways, and so determine the mass (density) function $P(Y = r)$.

$E[X_i] = (1-p) * 0 + p * 1 = p$
$E[X_i^2] = (1-p) * 0 + p * 1^2 = p$ and
$Var[X_i] = E[X_i^2] - E[X_i]^2 = p - p^2 = p(1-p)$

■

$$E[Y] = \sum E[X_i] = np$$

$$Var[Y] = \sum_{i=1}^{n} Var[X_i] = np(1-p)$$

■

$$P(Y = r) = \binom{n}{r} p^r (1-p)^{n-r}$$

■

5. 3-2

A contingency table with one row - why does chi-square work?

Suppose we make n independent Bernoulli trials, each with probability p.

A table of the results will look like this:

Success	Failure
x	$n - x$

A table of the expected results will look like this:

Success	Failure
np	$n - np$

The chi-square distribution with one degree of freedom has an expectation of 1. Show that the random variable that results from summing over both entries in the table also has an expectation of 1.

$$\chi^2 = \frac{(X-np)^2}{np} + \frac{(X-np)^2}{n(1-p)}$$

$$E[\chi^2] = \left[\frac{1}{np} + \frac{1}{n(1-p)}\right] E[(X-np)^2]$$

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

$$= \frac{Var[X]}{np(1-p)} = \frac{np(1-p)}{np(1-p)} = 1$$

■

6. 4-1

Poisson distribution as the limit of a binomial distribution

Random variable X_n has a binomial distribution with parameters n and p, expectation $\lambda = np$. So $p = \lambda/n$.

Its mass function is

$$P(X_n = x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!}\frac{\lambda^x}{n^x}(1-\frac{\lambda}{n})^{n-x}.$$

Take the limit as $n \to \infty$ to get the mass function for a Poisson random variable X:

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}.$$

$$\lim_{n\to\infty} \frac{n!}{x!\,(n-x)!}\frac{\lambda^x}{n^x}\left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$\lim_{n\to\infty} \frac{n!}{x!\,(n-x)!\,n^x} = \frac{n^x + O(n^{x-1})}{n^x k!} = \frac{1}{k!}$$

Where $O(n^{x-1})$ represents all other terms of order $x-1$ and below that will converge to 0 once we divide by n^x and take the limit

$$\left(1-\frac{\lambda}{n}\right)^{n-x} = \left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-x}$$

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{n} = e^{-\lambda}$$

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{-x} = 1$$

$$\lim_{n\to\infty} \frac{n!}{x!\,(n-x)!}\frac{\lambda^x}{n^x}\left(1-\frac{\lambda}{n}\right)^{n-x} = e^{-\lambda}\frac{\lambda^x}{x!}$$

■

7. 4-2

Prove that the Poisson distribution with parameter λ has mean and variance both equal to λ.

$$E[X] = \sum_{j=0}^{\infty} \frac{je^{-\lambda}\lambda^j}{j!}$$

$$= \sum_{j=1}^{\infty} \frac{je^{-\lambda}\lambda^j}{j!}$$

The sum when $j = 0$ is 0 so we can re-index to start at 1

$$= e^{-\lambda} \sum_{j=1}^{\infty} \frac{j\lambda\lambda^{j-1}}{j(j-1)!}$$

Factor out some things to cancel out the numerator and pull out a λ

$$= \lambda e^{-\lambda} \sum_{j=1}^{\infty} \frac{\lambda^{j-1}}{(j-1)!}$$

$$= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$$

Where we let $k = j - 1$

$$= \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Recognizing $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$ as the Taylor Series expansion for the exponential function

■

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

$$E[X^2] = \sum_{j=0}^{\infty} \frac{j^2e^{-\lambda}\lambda^j}{j!}$$

$$= \sum_{j=1}^{\infty} \frac{j^2e^{-\lambda}\lambda^j}{j!}$$

$$= e^{-\lambda} \sum_{j=1}^{\infty} \frac{j^2\lambda\lambda^{j-1}}{j(j-1)!}$$

$$= \lambda e^{-\lambda} \sum_{j=1}^{\infty} \frac{j\lambda^{j-1}}{(j-1)!}$$

$$= \lambda e^{-\lambda} \sum_{j=1}^{\infty} \frac{(j-1+1)\lambda^{j-1}}{(j-1)!}$$

$$= \lambda e^{-\lambda} \left(\sum_{j=1}^{\infty} \frac{(j-1)\lambda^{j-1}}{(j-1)!} + \sum_{j=1}^{\infty} \frac{\lambda^{j-1}}{(j-1)!} \right)$$

Splitting the summation up so we can again set up our Taylor expansion

$$= \lambda e^{-\lambda} \left(\lambda \sum_{j=2}^{\infty} \frac{\lambda^{j-2}}{(j-2)!} + \sum_{j=1}^{\infty} \frac{\lambda^{j-1}}{(j-1)!} \right)$$

$$= \lambda e^{-\lambda} \left(\lambda \sum_{i=0}^{\infty} \frac{\lambda^{i}}{i!} + \sum_{k=0}^{\infty} \frac{\lambda^{k}}{k!} \right)$$

$$= \lambda e^{-\lambda} \left(\lambda e^{\lambda} + e^{\lambda} \right) = \lambda^2 + \lambda$$

Using the $i = j - 2$ and $k = j - 1$ substitutions and Taylor expansion

$$Var[X] = \left(\lambda^2 + \lambda \right) - \lambda^2 = \lambda$$

■

8. 4-3

Prove that if X_1 and X_2 are independent Poisson random variables with parameters λ_1 and λ_2 respectively, then $X_1 + X_2$ is Poisson with parameter $\lambda_1 + \lambda_2$.

$$P(X_1 + X_2 = m) = \sum_{j=0}^{m} P(X_1 = j, X_2 = m - j)$$

$$= \sum_{j=0}^{m} P(X_1 = j)P(X_2 = m - j)$$

Using the independence assumption

$$= \sum_{j=0}^{m} \frac{e^{-\lambda_1}\lambda_1^j}{j!} \frac{e^{-\lambda_2}\lambda_2^{m-j}}{(m-j)!}$$

$$= e^{-(\lambda_1+\lambda_2)} \sum_{j=0}^{m} \frac{\lambda_1^j}{j!} \frac{\lambda_2^{m-j}}{(m-j)!}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{m!} \sum_{j=0}^{m} \frac{m!}{j!\,(m-j)!} \lambda_1^j \lambda_2^{m-j} = \frac{e^{-(\lambda_1+\lambda_2)}}{m!} (\lambda_1 + \lambda_2)^m$$

Inserting an $\frac{m!}{m!}$ to make the summation coincide with the Binomial Theorem

Thus, $X_1 + X_2$ is Poisson with parameter $\lambda_1 + \lambda_2$.

■

9. 5-1

(a) Let

$$I = \int_{-\infty}^{\infty} e^{-x^2/2} dx = \int_{-\infty}^{\infty} e^{-y^2/2} dy.$$

Being rather casual about infinite limits of integration, prove that $I^2 = 2\pi$.

(b) X has the standard normal distribution $N(0, 1)$ if its density function is

$$\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \text{ and from part (a), } \int_{-\infty}^{\infty} \rho(x) dx = 1.$$

Prove that $\mathrm{Var}[X] = 1$.

(c) Y has the normal distribution $N(\mu, \sigma)$ if its density function is

$$\rho(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{(y-\mu)^2/(2\sigma^2)}.$$

Prove that $E[Y] = \mu$ and $\mathrm{Var}[Y] = \sigma$.

$$I = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}}\, dx = \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}}\, dy$$

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}\, dx\, dy$$

Because I is equal to each both, when we square it, we can combine them into a double-integral

$$= \int_{r=0}^{\infty} \int_{\theta=0}^{2\pi} e^{-(r^2)/2}\, r\, dr\, d\theta$$

Converting to polar coordinates using $x^2 + y^2 \to r^2$ and $dx\, dy \to r\, dr\, d\theta$

$$= \int_{u=0}^{\infty} \int_{\theta=0}^{2\pi} e^{-u} du\, d\theta$$

Performing a u-substitution with $u = \frac{r^2}{2}$ and $du = r\, dr$

$$I^2 = 2\pi \int_{u=0}^{\infty} e^{-u} du = 2\pi$$

Pulling out the 2π because there is no dependency on θ

■

$$E[X] = \int_{-\infty}^{\infty} xe^{-\frac{x^2}{2}} dx = 0$$

By symmetry of the density function

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

$$E[X^2] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x * xe^{-\frac{x^2}{2}} dx$$

$\int u\, dv = uv - \int v\, du$ with $u = x$ and $du = dx$ and $dv = xe^{-x^2/2}dx$ and $v = -e^{-\frac{x^2}{2}}$

$$E[X^2] = xe^{-\frac{x^2}{2}}\Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = 0 + 1 = 1 = Var(X)$$

■

For N(μ, σ),

$$p(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

$$\frac{Y - \mu}{\sigma} = X$$

Where X is the standard normal distribution

$$E[Y] = E[\sigma X] + E[\mu] = 0 + \mu = \mu$$

With $\sigma E[X] = 0$

$$Var[Y] = Var[\sigma X] + Var[\mu] = \sigma^2 + 0 = \sigma^2$$

■

10. 5-2 Density function for the maximum

Suppose that $X_1, X_2, \cdots X_n$ are independent, identically distributed continuous random variables with density function f and distribution function F.

Let $X_{max} = \max\{X_1, X_2, \cdots X_n\}$.

Prove that the density function for X_{max} is $f_{max}(x) = nF(x)^{n-1}f(x)$.

Specialize to the case where the X_i are random variables from Unif$[0, \beta]$

$$F_{max}(X) = P(X_{max} \leq x) = P(X_1 \leq x, X_2 \leq x, \dots X_n \leq x)$$

$$= P(X_1 \leq x)P(X_2 \leq x) \dots P(X_n \leq x) = [F(x)]^n$$

$$f_{max}(x) = F'_{max}(x) = n[F(x)]^{n-1}F'(x) = n[F(x)]^{n-1}f(x)$$

■

$$f(x) = \frac{1}{\beta}$$

$$F(x) = \frac{x}{\beta}$$

$$f_{max}(x) = n\left(\frac{x}{\beta}\right)^{n-1}\frac{1}{\beta} = \frac{nx^{n-1}}{\beta}$$

■

11. 5-3

For random variable X, define the moment generating function

$$M(t) = E[e^{tX}].$$

Prove that

- The nth derivative of $M(t)$, evaluated at $t = 0$, is equal to the nth moment $E[X^n]$.
- If X_1 and X_2 are independent random variables with moment generating functions $M_1(t)$ and $M_2(t)$, then

$$M_{X_1+X_2}(t) = M_1(t)M_2(t)$$

$$e^{tX} = 1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots$$

$$M(t) = E\left[1 + tX + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots\right] = 1 + tE[X] + \frac{t^2}{2!}E[X^2] + \frac{t^3}{3!}E[X^3] + \cdots$$

$$M'(t) = 0 + E[X] + tE[X^2] + \frac{t^2}{2!}E[X^3] + \frac{t^3}{3!}E[X^4] + \cdots$$

$$M'(0) = E[X]$$

$$M''(t) = 0 + 0 + E[X^2] + tE[X^3] + \frac{t^2}{2!}E[X^4] + \cdots$$

$$M''(0) = E[X^2]$$

Proceeding along the same logic,

$$M^{(n)}(0) = E[X^n]$$

■

$$M_{X_1+X_2}(t) = E\left[e^{t[X_1+X_2]}\right] = E[e^{tX_1}e^{tX_2}] = E[e^{tX_1}]E[e^{tX_2}] = M_{X_1}(t)M_{X_2}(t)$$

■

12. 6-1 Confidence intervals as random variables

Suppose that we are drawing samples of size n from a known population with mean μ and variance σ^2. The sample mean $\overline{X}$ is a random variable whose expectation is also μ. Let α be a smallish number, typically $\alpha = 0.05$. Then a "$1-\alpha$ confidence interval" is specified by two random variables L and U with the property that

$$P(L \geq \mu) = P(U \leq \mu) = \alpha/2.$$

Thus the probability of the event $L < \mu < U$ is $1 - \alpha$ (typically 95%).

Let q_1 and q_2 denote the $\alpha/2$ and $(1-\alpha/2)$ quantiles for the sampling distribution of X.

(a) Show that $U = \overline{X} + \mu - q_1$ and that $L = \overline{X} + \mu - q_2$. Explain why it is reasonable for U to depend on the "lower" quantile q_1 and vice versa.

(b) Show that if the sampling distribution is symmetrical about μ, then $U = \overline{X} + q_2 - \mu$ and that $L = \overline{X} - (\mu - q_1)$.

(c) For the normal distribution $N(0,1)$ and $\alpha = .05$, $\mu = 0$, $q_1 = -1.96$ and $q_2 = 1.96$. Show that if the central limit theorem applies, then $L = \overline{X} - 1.96\sigma/\sqrt{n}$; $U = \overline{X} + 1.96\sigma/\sqrt{n}$.

$$P(U < \mu) = P(\bar{X} + \mu - q_1 \leq \mu) = P(\bar{X} \leq q_1) = \frac{\alpha}{2}$$

$$P(L \geq \mu) = P(\bar{X} + \mu - q_2 \geq \mu) = P(\bar{X} \geq q_2) = \frac{\alpha}{2}$$

U depends on the lower quantile and L the upper quantile because U is the upper bound for the lower $\left(1 - \frac{\alpha}{2}\right)$ portion and L is the lower bound for the upper portion.

■

For a symmetrical distribution about μ, $q_1 + q_2 = 2\mu$

$$U = \bar{X} + \mu - q_1 = \bar{X} + \mu + q_2 - 2\mu += \bar{X} + q_2 - \mu$$

$$L = \bar{X} + \mu - q_2 = \bar{X} + \mu + q_1 - 2\mu += \bar{X} + q_1 - \mu$$

■

$$U = \bar{X} + \mu + \frac{1.96\sigma}{\sqrt{n}} - \mu = \bar{X} + \frac{1.96\sigma}{\sqrt{n}}$$

$$L = \bar{X} + \mu - \frac{1.96\sigma}{\sqrt{n}} - \mu = \bar{X} - \frac{1.96\sigma}{\sqrt{n}}$$

13. 6-2 (Shortened)

A normal random variable $X \sim N(\mu, \sigma^2)$ has density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

The moment generating function of X is

$$M(t) = e^{\mu t + \sigma^2 t^2/2}.$$

Prove the following, using the moment generating function.

- $E[X] = \mu$.
- $\mathrm{Var}[X] = \sigma^2$.
- If $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma^2)$, and X_1 and X_2 are independent, then $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$M'(t) = (\mu + \sigma^2 t) e^{\mu t + \frac{\sigma^2 t^2}{2}}$

$M'(0) = E[X] = \mu$

■

$M''(t) = \sigma^2 e^{\mu t + \frac{\sigma^2 t^2}{2}} + (\mu + \sigma^2 t) e^{\mu t + \frac{\sigma^2 t^2}{2}}$

$M''(0) = \sigma^2 + \mu^2$

$Var[X] = \sigma^2 + \mu^2 - \mu^2 = 0$

■

$M_{X_1+X_2}(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}} e^{\mu_2 t + \frac{\sigma_2^2 t^2}{2}}$

$= e^{(\mu_1+\mu_2)t + \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}$

$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

■

14. 7-1

Prove that if $x_1, x_2, \cdots x_n$ are an independent random sample from a normal distribution with unknown parameters μ and σ, the maximum-likelihood estimators of the parameters are

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$L(\mu, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\ln L(\mu, \sigma) = \sum_{i=1}^{n} -\ln \sigma - \frac{\ln 2\pi}{2} - \frac{(\mu - x_i)^2}{2\sigma^2}$$

Taking the log-likelihood converts the product into a sum and we can use the symmetry property

$$\frac{(x_i - \mu)^2}{2\sigma^2} = \frac{(\mu - x_i)^2}{2\sigma^2}$$

To optimize and find the MLE, we take the partial derivative with respect to each variable

$$\frac{d}{d\mu}\ln L = \sum_{i=1}^{n} -\frac{2(\mu - x_i)}{2\sigma^2} = 0$$

$$-n\mu + \sum_{i=1}^{n} x_i = 0$$

$$n\mu = \sum_{i=1}^{n} x_i$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

■

$$\frac{d}{d\sigma}\ln L = \sum_{i=1}^{n} -\frac{1}{\sigma} - 0 - \sum_{i=1}^{n} \frac{-2(x_i - \mu)^2}{2\sigma^3} = 0$$

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Substituting in $\bar{x}$ for the estimate of μ

■

15. 7-2

Suppose that $x_1, x_2, \cdots x_n$ are a random sample from a uniform distribution Unif$[0, \beta]$ with unknown parameter β. Show that the maximum likelihood estimator of β is $\max x_i$ and that this estimator can be made unbiased by multiplying it by $(n+1)/n$.

$$f(x) = \frac{1}{\beta}$$

$$L(x) = \prod_{i=1}^{n} \frac{1}{\beta} I(x_1, x_2, \dots x_n)$$

Where $I(n)$ is the indicator function that is 1 if $0 \leq x \leq \beta$ and 0 otherwise

$L(x) = \frac{1}{\beta^n}$ if $I(x_1, x_2, \dots x_n) = \max(x_1, x_2, \dots x_n)$

$I(X_1, X_2, \dots, X_n) = 0$ if any one of the factors is 0 (falls outside the allowed interval), so the only way that will not happen is if we take the maximum of the observations.

MLE: $\hat{\beta} = \max x_i$

■

$$f(x) = n\left(\frac{x}{\beta}\right)^{n-1}\frac{1}{\beta}$$

$$E[\max(x_i)] = \int_0^{\beta} x \frac{nx^{n-1}}{\beta^n} dx$$

$$= \frac{n}{n+1}\frac{\beta^{n+1}}{\beta^n} = \frac{n}{n+1}\beta$$

$$\hat{\beta}_{unbiased} = \frac{(n+1)}{n} \max x_i$$

■

16. 7-3

Suppose that $X_1, X_2, \cdots X_n$ are independent random variables, all with expectation 0, variance 1, and the same moment generating function $M(t)$. Define $Z_n = \frac{1}{\sqrt{n}}(X_1 + X_2 + \cdots + X_n)$,
and call its moment generating function $H_n(t)$.
Using the fact that if $\lim_{n\to\infty} nr_n = 0$ then

$$\lim_{n\to\infty}(1 + \frac{x}{n} + r_n)^n = e^x$$

prove that

$$\lim_{n\to\infty} H_n(t) = e^{\frac{t^2}{2}},$$

the moment-generating function of the standard normal distribution.

$$M(t) = M(0) + M'(0)t + \frac{M''(0)t^2}{2!} + \frac{M'''(0)t^3}{3!} + \cdots$$

Call the remainder starting from the 3rd derivative $r(t)$

$$\lim_{t\to 0}\frac{r(t)}{t^2} = 0$$

$$M(0) = 1$$

$$M'(0) = E[X] = 0$$

$$M''(0) = E[X^2] = 1$$

$$M(t) = 1 + 0t + \frac{t^2}{2} + r(t)$$

$$H_n(t) = \left(1 + \frac{t^2}{2n} + \frac{M_3 t^3}{6n^{1.5}} + r(t)\right)^n$$

Dividing the random variable by $\sqrt{n}$ means consecutive term gets divided by a higher power of n

Let $x = \frac{t^2}{2}$

$$\lim_{n\to\infty}\left(1 + \frac{x}{n} + r(n)\right)^n = e^X = e^{\frac{t^2}{2}}$$

■

17. 8-1 Define the gamma function for $r > 0$ by $\Gamma(r) = \int_0^\infty x^{r-1} e^{-x} dx$. Prove that

- $\Gamma(r+1) = r\Gamma(r)$ if $r > 0$.
- For integer $n > 0, \Gamma(n+1) = n!$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

$$\Gamma(r+1) = \int_0^\infty x^r e^{-x} dx$$

$$u = x^r, v = e^{-x}$$

$$du = rx^{r-1}, dv = -e^{-x}$$

$$= x^r e^{-x}|_0^\infty + r\int_0^\infty x^{r-1} e^{-x} dx = 0 + r\Gamma(r) = r\Gamma(r)$$

■

$$\Gamma(n+1) = n\Gamma(n)$$

$$= n(n-1)\Gamma(n-1)$$

$$= \frac{n(n-1) * \dots 2}{n!}\Gamma(1) = n!$$

■

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty x^{-\frac{1}{2}} e^{-x} dx$$

$$x = \frac{u^2}{2}$$

$$dx = u\, du$$

$$= \int_0^\infty \frac{\sqrt{2}}{u} e^{-\frac{u^2}{2}} udu$$

$$= \frac{1}{\sqrt{2}} \int_{-\infty}^\infty e^{-\frac{u^2}{2}} udu$$

$$= \frac{1}{\sqrt{2}} \sqrt{2\pi} = \sqrt{\pi}$$

■

18. 8-2 A random variable X has the gamma distribution if its probability density function is

$$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}, x \geq 0$$

Prove the following:

- The moment generating function is $M(t) = (\frac{\lambda}{\lambda - t})^r$.

- $$E[X] = \frac{r}{\lambda}; E[X^2] = \frac{(r+1)r}{\lambda^2}; \mathrm{Var}[X] = \frac{r}{\lambda^2}$$

- If $X_1, X_2, \cdots X_n$ are independent random variables with $X_i \sim \mathrm{Gamma}(r_i, \lambda)$, then

$$X = X_1 + X_2 + \cdots + X_n \sim \mathrm{Gamma}(r_1 + r_2 + \cdots + r_n, \lambda).$$

$$M(t) = E[e^{tX}] = \int_0^\infty f(x) e^{tx} dx$$

$$= \frac{\lambda^r}{\Gamma(r)} \int_0^\infty x^{r-1} e^{-(\lambda - t)x} dx$$

$$x = \frac{u}{\lambda - t}$$

$$dx = \frac{du}{\lambda - t}$$

$$= \frac{\lambda^r}{\Gamma(r)} * \frac{1}{(\lambda - t)^r} \int_0^\infty u^{r-1} e^{-u} dx$$

$$\int_0^\infty u^{r-1} e^{-u} dx = \Gamma(r)$$

$$= \frac{\lambda^r}{(\lambda - t)^r}$$

■

$$M(t) = \frac{\lambda^r}{(\lambda - t)^r}$$

$$M'(t) = \frac{r\lambda^r}{(\lambda - t)^{r+1}}$$

$$M''(t) = \frac{r(r+1)\lambda^r}{(\lambda - t)^{r+2}}$$

$$E[X] = M'(0) = \frac{r}{\lambda}$$

$$E[X^2] = M''(0) = \frac{r(r+1)}{\lambda^2}$$

$$Var[X] = E[X^2] - E[X]^2 = \frac{r(r+1)}{\lambda^2} - \frac{r^2}{\lambda^2} = \frac{r}{\lambda^2}$$

■

$$M_X(t) = M_{X_1}(t) M_{X_2}(t) * \dots M_{X_n}(t) = \frac{\lambda^{r_1}}{(\lambda - t)^{r_1}} \frac{\lambda^{r_2}}{(\lambda - t)^{r_2}} \cdots \frac{\lambda^{r_n}}{(\lambda - t)^{r_n}} = \left(\frac{\lambda}{\lambda - t}\right)^{r_1 + r_2 + \cdots r_n}$$

$$X \sim Gamma(r_1 + r_2 + \cdots r_n, \lambda)$$

■

19. 8-3 A random variable X has the chi-square distribution with m degrees of freedom if

$$X \sim \chi_m^2 \sim \text{Gamma}(\frac{m}{2}, \frac{1}{2})$$

. Using properties of the gamma distribution, prove the following:

- If $X_1, X_2, \cdots X_n$ are independent chi-square random variables with degrees of freedom $m_1, m_2, \cdots m_n$, then $X = X_1 + X_2 + \cdots + X_n$ is chi-square with $m = m_1 + m_2 + \cdots + m_n$ degrees of freedom.
- If $Z \sim N(0,1)$, then Z^2 is chi-square with one degree of freedom.
- If $Z_1, Z_2, \cdots Z_k$ are independent $N(0,1)$ random variables, then $X = Z_1^2 + Z_2^2 + \cdots Z_k^2$ has a chi-square distribution with k degrees of freedom.
- The moment generating function of X is $M(t) = (1-2t)^{-k/2}$.

$$X = X_1 + \cdots X_n \sim Gamma(r_1 + r_2 + \cdots r_n, \lambda)$$

$$X \sim Gamma\left(\frac{m}{2}, \frac{1}{2}\right) \sim \chi_m^2$$

■

$$\chi_1^2 \sim Gamma(\frac{1}{2}, \frac{1}{2})$$

$$f(x) = \frac{\frac{1}{2}^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)} x^{-\frac{1}{2}} e^{-\frac{1}{2}x} = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}$$

Need distribution function, and then take the derivative to get the density function

$$F_{Z^2}(x) = P(Z^2 \leq x) = P(-\sqrt{x} \leq Z \leq \sqrt{x})$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-\frac{z^2}{2}} dz$$

$$f_{Z^2}(x) = F'(x) = \frac{1}{\sqrt{2\pi}} \left[\frac{1}{2\sqrt{x}} e^{-\frac{x}{2}} + \frac{1}{2\sqrt{x}} e^{-\frac{x}{2}} \right]$$

$$= \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}$$

Equivalent to $f(x)$ from above, so $Z^2 \sim \chi_1^2$

■

$m = m_1 + m_2 + \cdots m_n$ degrees of freedom

The sum is obvious since we sum 1+1+1 k times so $m = k$ degrees of freedom.

■

$$r = \frac{k}{2}$$

$$\lambda = \frac{1}{2}$$

$$M(t) = \left(\frac{\frac{1}{2}}{\frac{1}{2} - t}\right)^{\frac{k}{2}}$$

$$= \left(\frac{1}{1 - 2t}\right)^{\frac{k}{2}}$$

$$= (1 - 2t)^{-\frac{k}{2}}$$

■

20. 8-4 Let X_1, X_2 be a random sample from $N(0,1)$, with sample mean $\overline{X}$ and sample variance S^2. Prove the following:

- $X_1^2 + X_2^2 = 2\overline{X}^2 + S^2$.
- $E[\overline{X}^2 S^2] = E[\overline{X}^2]E[S^2]$
  (Use the fact that if X is standard normal, $E[X^2] = 1$ and $E[X^4] = 3$.

$$\bar{X} = \frac{X_1 + X_2}{2}$$

$$S^2 = \frac{1}{2-1}[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2]$$

$$2\bar{X}^2 = \frac{(X_1 + X_2)^2}{2}$$

$$S^2 = \left[\frac{(X_1 - X_2)^2}{2} - \frac{(X_2 - X_1)^2}{2}\right] = \frac{(X_2 - X_1)^2}{2}$$

$$2\bar{X}^2 + S^2 = \frac{(X_1 + X_2)^2}{2} + \frac{(X_1 - X_2)^2}{2} = X_1^2 + X_2^2$$

$$X_1^2 + X_2^2 = 2\bar{X}^2 + S^2$$

■

$$E[\bar{X}^2 S^2] =$$

$$E\left[\frac{(X_1 + X_2)^2}{4} * \frac{(X_1 - X_2)^2}{2}\right] = \frac{1}{8}E[(X_1^2 - X_2^2)^2]$$

$$= \frac{1}{8}(E[(X_1^4) - 2E[X_1^2 X_2^2] + E[X_2^4])$$

$$= \frac{1}{8}[3 - 2 * 1 * 1 + 3] = \frac{1}{2}$$

$$E[\bar{X}^2] = \frac{1}{2} = Var(\bar{X})$$

$$E[S^2] = \frac{1}{2}(E(X_1^2 - 2E[X_1]E[X_2] + E[X_2^2])$$

$$= \frac{1}{2}[1 - 0 + 1] = 1$$

$$E[\bar{X}^2 S^2] = E[\bar{X}^2]E[S^2]$$

■

21. 8-5 Let $X_1, X_2, \cdots X_n$ be a random sample from $N(\mu, \sigma^2)$, with sample mean $\overline{X}$ and sample variance S^2. It continues to be true that $\overline{X}$ and S^2 are independent random variables. Define

$$U = \frac{1}{\sigma^2}\sum_{i=1}^{k}(X_i - \mu)^2; V = \frac{1}{\sigma^2}\sum_{i=1}^{k}(X_i - \overline{X})^2; W = \frac{1}{\sigma^2}n(\overline{X} - \mu)^2.$$

Prove the following:

- $U = V + W$.
- $(n-1)S^2/\sigma^2$ has a chi-square distribution with $n-1$ degrees of freedom.

$$U = \frac{1}{\sigma^2}\sum_{i=1}^{k}(X_i - \mu)^2$$

$$\sum_{i=1}^{k}(X_i - \mu)^2 = \sum_{i=1}^{k}(X_i - \bar{X} + \bar{X} - \mu)^2$$

$$= \sum_{i=1}^{k}(X_i - \bar{X})\text{^}2 + (\bar{X} - \mu)\sum_{i=1}^{k}(X_i - \bar{X}) + n(\bar{X}_i - \mu)^2$$

$$\sum_{i=1}^{k}(X_i - \bar{X}) = 0$$

$$U = \frac{1}{\sigma^2}\sum_{i=1}^{k}(X_i - \bar{X})^2 + \frac{1}{\sigma^2}n(\bar{X}_i - \mu)^2 = V + W$$

■

I made this part up but it seems reasonable to me!

U is χ^2 with n degrees of freedom

W is χ^2 with 1 degree of freedom

So the remaining V is χ^2 with n-1 degrees of freedom (from independence) and

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

■

22. 8-6 (Greatly shortened)

Let $Z \sim N(0,1)$ and let W denote a chi-square distribution with k degrees of freedom, independent of Z. Let T be the ratio

$$T = \frac{Z}{\sqrt{W/k}}$$

Let $f_W(w)$ denote the density function for the chi-square distribution; let $f_Z(z)$ denote the density function for the normal distribution.

Prove that the density function for T is

$$f_T(t) = \int_0^\infty \sqrt{\frac{w}{k}} f_Z(t\sqrt{\frac{w}{k}}) f_W(w) dw.$$

You can stop here, not doing the messy algebra that converts this to

$$f_T(t) = \frac{\Gamma((k+1)/2)}{\Gamma(k/2)\sqrt{k\pi}}(1+\frac{t^2}{k})^{-(k+1)/2}.$$

$$F_T(t) = P(T \le t) = P\left(\frac{Z}{\sqrt{\frac{W}{k}}} \le t\right)$$

$$= P\left(Z \le \sqrt{\frac{W}{k}}t\right)$$

$$= \int_0^\infty f_w(w)\left(\int_{-\infty}^{\sqrt{\frac{W}{k}}t} f_z(z)dz\right)dw$$

$$f(t) = F_T'(t) = \int_0^\infty f_w(w)\sqrt{\frac{W}{k}} * \frac{1}{\sqrt{2\pi}} e^{-\frac{w}{k}*\frac{t^2}{2}} dw$$

$$= \frac{\left(\frac{1}{2}\right)^{\frac{k}{2}}}{\Gamma\left(\frac{k}{2}\right)} \int_0^\infty w^{\frac{k}{2}-1} e^{-\frac{1}{2}w} \sqrt{\frac{W}{k}} \frac{1}{\sqrt{2\pi}} e^{-w*\frac{t^2}{2k}} dw$$

$$= \int_0^\infty \sqrt{\frac{W}{k}} f_Z\left(t\sqrt{\frac{W}{k}}\right) f_W(w) dw$$

Messy algebra converts it to:

$$= \frac{\Gamma\left(\frac{\mathrm{k}+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{k\pi}}\left(1+\frac{x^2}{k}\right)^{-\left(\frac{k+1}{2}\right)}$$

■

23. 8-7 Prove that if $X_1, X_2, \cdots X_n$ are a random sample from $N(\mu, \sigma^2)$, then

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a t distribution with $n-1$ degrees of freedom.

X is normal so $\bar{X}$ is normal, and we subtract mean, so we are left with something with 0 mean.

$$\bar{X} - \mu \sim N(0, \frac{\sigma}{\sqrt{n}})$$

$$\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu) \sim N(0,1)$$

Someone please check the rest for me?

$$W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

$$T = \frac{\frac{\sqrt{n}}{\sigma}(\bar{X} - \mu)}{\frac{S\sqrt{n-1}}{\sigma} / \sqrt{n-1}}$$

Needed to divide W by $\sqrt{n-1}$ too

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

■

24. 8-8 Student t confidence interval

Let $X_1, X_2, \cdots X_n$ be a random sample from $N(\mu, \sigma^2)$, with both μ and σ unknown. The sample mean is $\overline{X}$; the sample variance is S^2.

Let q denote the $(1-\alpha/2)$ quantile of the Student t distribution with $n-1$ degrees of freedom. By symmetry, $-q$ is the $\alpha/2$ quantile.

Define random variables

$$L = \overline{X} - \frac{qS}{\sqrt{n}}; U = \overline{X} + \frac{qS}{\sqrt{n}}.$$

Prove that $P(L > \mu) = \alpha/2$ and $P(U < \mu) = \alpha/2$,
so that $P(L \leq \mu \leq U) = 1 - \alpha$. and $[L, U]$ is a $1-\alpha$ confidence interval.

$$\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

$$P\left(\frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \leq q\right) = 1 - \frac{\alpha}{2}$$

By definition of q

$$P\left(\bar{X} - \mu \leq \frac{qS}{\sqrt{n}}\right) = 1 - \frac{\alpha}{2}$$

$$P\left(\bar{X} - \mu > \frac{qS}{\sqrt{n}}\right) = \frac{\alpha}{2}$$

$$P\left(\bar{X} - \frac{qS}{\sqrt{n}} > \mu\right) = \frac{\alpha}{2}$$

$$P(L > \mu) = \frac{\alpha}{2}$$

■

By symmetry,

$$P\left(\bar{X} - \mu < \frac{-qS}{\sqrt{n}}\right) = \frac{\alpha}{2}$$

$$P\left(\bar{X} + \frac{qS}{\sqrt{n}} < \mu\right) = \frac{\alpha}{2}$$

$$P(U < \mu) = \frac{\alpha}{2}$$

■

$$P(L \leq \mu \leq U) = 1 - \frac{\alpha}{2} - \frac{\alpha}{2} = 1 - \alpha$$

■

25. 9-1 Covariance and correlation

(a) The covariance of random variables X and Y is defined as
$\mathrm{Cov}[X,Y] = E[(X-\mu_X)(Y-\mu_Y)]$
Prove that $\mathrm{Cov}[X,Y] = E[XY] - E[X]E[Y]$.

(b) The correlation coefficient of random variables X and Y is defined as

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

Prove that $|\rho(X,Y)| \le 1$.

(c) Prove that when calculating the sample correlation r, you can divide $\sum(x_i-\overline{x})(y_i-\overline{y})$ by $n, n-1$, or 1 in the numerator, as long as you do the same thing in the denominator.

$Cov[X,Y] = E[(X-\mu_X)(Y-\mu_Y)]$

$= E[XY] - \mu_X E[Y] - \mu_Y E[X] + \mu_X \mu_Y$

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

■

$$Z_X = \frac{X-\mu_X}{\sigma_X}$$

$$Z_Y = \frac{Y-\mu_Y}{\sigma_Y}$$

Both have expectation 0 and variance 1 but are correlated

$Var[Z_X \pm Z_Y] = E[Z_X^2] + E[Z_Y^2] \pm 2E[Z_X Z_Y]$

$Var[Z_X \pm Z_Y] = 1 + 1 \pm 2\rho(X,Y) \ge 0$

$\pm\rho(X,Y) \le 1$

$|\rho(X,Y)| \le 1$

■

$$r = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2}}$$

The $\frac{1}{n-1}$ or any constant there cancels out, so is unnecessary

26. 9-2 Least-squares regression

You have values x_i of a "predictor" and matching values y_i of a "response." Your goal is to minimize the sum of squares of the prediction errors,

$$g(a,b) = \sum_{i=1}^{n} (a + bx_i - y_i)^2.$$

Prove that

$$b = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}, a = \overline{y} - b\overline{x}.$$

$$\frac{\partial g}{\partial a} = 2\sum_{i=1}^{n}(a + bx_i - y_i) = 0$$

$$na - nb\bar{x} - n\bar{y} = 0$$

$$\bar{y} = a + b\bar{x}$$

$$\frac{\partial g}{\partial b} = 2\sum_{i=1}^{n} x_i(a + bx_i - y_i) = 0$$

$$\sum_{i=1}^{n} \bar{x}(a + bx_i - y_i) = 0$$

Now we subtract the prior 2 equations

$$\sum_{i=1}^{n}(x_i - \bar{x})(\bar{y} - b\,\bar{x} + bx_i - y_i) = 0$$

$$b\sum_{i=1}^{n}(x_i - \bar{x})^2 - \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 0$$

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

■

27. 9-3 The connection between correlation and the slope of the regression line.

Define $ss_{xy} = \sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})$; $ss_x = \sum_{i=1}^{n}(x_i - \overline{x})^2$; $ss_y = \sum_{i=1}^{n}(y_i - \overline{y})^2$.

Prove that $r^2 ss_y = b^2 ss_x$.

An observation is y_i; a predicted observation is $\hat{y}_i = a + bx_i; \overline{y} = a + b\overline{x}$. Prove that the ratio of the variance of the predicted y's to the variance of the observed y's equals R-squared, the square of the sample correlation r.

Prove that the ratio of the variance of the residuals $y - \hat{y}$ to the variance of the observed y's equals $1 - r^2$.

$$r^2 ss_y = \frac{ss_{xy}^2}{ss_x ss_y} ss_y = \frac{ss_{xy}^2}{ss_x}$$

$$b^2 ss_x = \frac{ss_{xy}^2}{ss_x^2} ss_x = \frac{ss_{xy}^2}{ss_x}$$

$$r^2 ss_y = b^2 ss_x$$

■

$$\frac{Pred\ Var}{Observed\ Var} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$\hat{y}_i = a + bx_i$$

$$b = \sqrt{\frac{ss_y}{ss_x}} = r$$

$$= \frac{\sum_{i=1}^{n}(a + bx_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$= \frac{\sum_{i=1}^{n} b^2(x_i - \bar{x})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$= \frac{b^2 ss_x}{ss_y} = r^2$$

■

$$\frac{Residual\ Var}{Observed\ Var} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$= \frac{\sum_{i=1}^{n}(y_i - a - bx_i\ )^2}{ss_y}$$

$$= \frac{\sum_{i=1}^{n}(y_i - \bar{y} - bx_i - b\bar{x})^2}{ss_y}$$

$$= \frac{\sum(y_i - \bar{y})^2 - \sum 2b(y_i - \bar{y})(x_i - \bar{x}) + b^2\sum(x_i - \bar{x})^2)}{ss_y}$$

$$= \frac{ss_y - 2b^2 ss_x + b^2 ss_x}{ss_y}$$

$$= \frac{ss_y - b^2 ss_x}{ss_y}$$

$$= \frac{ss_y - r^2 ss_y}{ss_y} = 1 - r^2$$

∎

28. 9-4 Maximum likelihood regression

You have a fixed set of values, x_i, of a "predictor" variable.

For each x_i, the response Y_i is a random variable whose expectation is $\mu_i = \alpha + \beta x_i$ and whose variance is σ^2. The residuals $Y_i - \mu_i$ are independent.

Given a set of pairs of values $(x_1, Y_1), (x_2, Y_2), \cdots (x_n, Y_n)$, prove that the maximum-likelihood estimates of α and β sastify the equations

$$\sum_{i=1}^{n}(\hat{\alpha} - \hat{\beta}x_i - Y_i) = 0.$$

$$\sum_{i=1}^{n} x_i(\hat{\alpha} - \hat{\beta}x_i - Y_i) = 0.$$

and that

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i-1}^{n}(Y_i - \hat{\alpha} - \hat{\beta}x_i)^2.$$

You do not need to redo the algebra(from 9-2) to show that

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}, \hat{\alpha} = \overline{Y} - \hat{\beta}\overline{x}.$$

$$P(Y_1, Y_2, \dots Y_n) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(Y_i - \alpha - \beta x_i)^2}{2\sigma^2}}$$

$$-\ln P = \sum_{i=1}^{n} \ln\sigma + \ln\sqrt{2\pi} + \frac{(\alpha + \beta x_i - Y_i)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial\alpha} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(\alpha + \beta x_i - Y_i) = 0$$

$$\frac{\partial}{\partial\beta} = \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i(\alpha + \beta x_i - Y_i) = 0$$

$$\frac{\partial}{\partial\sigma} = \frac{n}{\sigma} - \frac{1}{\sigma^3}\sum_{i=1}^{n}(Y_i - \alpha - \beta x_i)^2 = 0$$

What's the proof of this last piece?

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \left(Y_i - \hat{\alpha} - \hat{\beta} x_i\right)^2}{n}$$

INCOMPLETE

■

29. 9-5 Logistic regression

You have a fixed set of values, x_i, of a "predictor" variable. Each "response" variable Y_i is a Bernoulli random variable with parameter p_i.

Assume that

$$p_i = \frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}.$$

(a) Prove that $\alpha + \beta x_i$ is equal to the "log odds" $\ln \frac{p_i}{1-p_i}$.

(b) Prove that $0 < p_i < 1$.

(c) Given a set of pairs of values $(x_1, Y_1), (x_1, Y_1), \cdots (x_n, Y_n)$, form the likelihood function $L(\alpha, \beta)$ and express its logarithm in terms of α and β. Do not attempt to maximize it!

$$\frac{p_i}{1-p_i} = \frac{\frac{e^{\alpha+\beta x_i}}{1+e^{\alpha+\beta x_i}}}{\frac{1}{1+e^{\alpha+\beta x_i}}} = e^{\alpha+\beta x_i}$$

$$\ln \frac{p_i}{1-p_i} = \alpha + \beta x_i$$

■

Let $x = \alpha + \beta x_i$

$$\lim_{x\to-\infty} \frac{e^x}{1+e^x} = 0$$

$$\lim_{x\to\infty} \frac{e^x}{1+e^x} = 1$$

Since p_i is an increasing function, 0 < p_i <1.

■

$$L(\alpha, \beta) = \prod_{i=1}^{n} p_i^{Y_i}(1-p_i)^{1-Y_i}$$

$$\ln L(\alpha, \beta) = \sum_{i=1}^{n} Y_i \ln p_i + (1-Y_i)\ln(1-p_i)$$

$$= \sum_{i=1}^{n} Y_i(\alpha+\beta x_i) - \ln\left(1+e^{\alpha+\beta x_i}\right)$$

■

30. 10-1

(a) State and prove Bayes' theorem for two events A and B, neither of which has probability zero.

(b) Specialize to the case where X are some data (one or more random variables) from a discrete probability distribution that is specified by one or more parameters collectively represented by θ. The event A is "the parameter value is θ_j"; the event B is "the observed data are X."

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

Proof by Tautology

■

$$P(\theta = \theta_j | X) = \frac{P(\theta = \theta_j)P(X|\theta = \theta_j)}{P(X)} = \frac{P(\theta = \theta_j)P(X|\theta = \theta_j)}{\sum_{i=1}^{n} P(\theta_j)\, P(X|\theta = \theta_j)}$$

$P(\theta = \theta_j)$ is the Bayesian prior

$P(X|\theta = \theta_j)$ is the likelihood

$P(\theta = \theta_j|X)$ posterior distribution

■

31. 10-2

Suppose that random variable X has a beta distribution, i.e. its probability density function is

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 \leq x \leq 1$$

Knowing that $\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx = 1$, prove that

$$E[X] = \frac{\alpha}{\alpha+\beta}$$

Credit to Joe Palin for explaining to me the missing steps!

$$E[X] = \int_0^1 x f(x) dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1} dx$$

$$\int_0^1 x^{\alpha}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)}$$

From recognition that

$$\frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)} \int_0^1 x^{\alpha+1-1}(1-x)^{\beta-1} dx = 1$$

$$E[X] = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+1)}$$

$$\frac{\Gamma(x+1)}{\Gamma(x)} = x$$

$$E[X] = \frac{\alpha}{\alpha+\beta}$$

■

32. 10-3

Suppose that the parameter θ for a binomial distribution has the prior distribution $\theta \sim$ Beta(α, β) and the data are given by the binomial distribution $X \sim$ Binom(n, θ). Prove that the posterior distribution $\theta|X \sim$ Beta$(\alpha + x, \beta + n - x)$.

$$\text{Reminder: } \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} d\theta = 1.$$

$$p(\theta|x) = \frac{\pi(\theta) f(x|\theta)}{\int_{-\infty}^{\infty} \pi(\theta) f(x|\theta) d\theta}$$

$$\pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$\propto$ means is proportional to

$$f(x|\theta) \propto \theta^{x}(1-\theta)^{n-x}$$

$$p(\theta|x) = \pi(\theta) f(x|\theta) d\theta \propto \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}$$

$$p(\theta|x) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}$$

$$p(\theta|x) \sim Beta(\alpha+x, \beta+n-x)$$

■

33. 11-1

Pivotal statistic – location parameter

Suppose that $X_1, X_2, \cdots X_n$ are random variables from a distribution with parameter θ. A *pivotal statistic* is a function $h(X_1, X_2, \cdots X_n, \theta)$ whose distribution does not depend on θ or on any parameters with unknown values. The parameter θ is called a *location parameter* if the distribution of $X - \theta$ does not depend on θ.

(a) Show that for a normal distribution with known σ, μ is a location parameter and that given one sample X from this distribution, there are random variables L and U, with values that depend on X but not on θ, such that $P(\mu < L) = 0.025$ and $P(\mu > U) = 0.025$

(b) Show that if $X \sim \text{Unif}(\theta - 1, \theta + 1)$, then θ is a location parameter, and explain how to determine L and R for a 90% confidence interval.

$(X-\mu) \sim N(0, \sigma^2)$ does not depend on μ, so is a location parameter.

$P(q_{0.025} < X - \mu < q_{0.975}) = 0.95$

$P(q_{0.025} < X - \mu) = P(X - \mu < q_{0.975}) = 0.975$

$P(q_{0.025} > X - \mu) = P(X - \mu > q_{0.975}) = 0.025$

$P(X - q_{0.025} < \mu) = P(\mu < X - q_{0.975}) = 0.025$

$P(U < \mu) = P(\mu < L) = 0.025$

$U = X - q_{0.025}, L = X - q_{0.975}$

■

$(X-\theta) \sim U(-1,1)$ does not depend on θ, so is a location parameter.

$P(q_{.05} < X - \theta < q_{.95}) = .90$

$P(q_{.05} < X - \theta) = P(X - \theta < q_{.95}) = .95$

$P(q_{.05} > X - \theta) = P(X - \theta > q_{.95}) = .05$

$P(X - q_{.05} < \theta) = P(\theta < X - q_{.95}) = .05$

$P(U < \theta) = P(\theta < L) = .05$

$L = X - q_{.95} = X - 0.9, U = X - q_{.05} = X + 0.9$

34. 11-2

Pivotal statistic – scale parameter

A parameter θ is called a *scale parameter* if the distribution of X/θ does not depend on θ.

(a) Show that if X has the distribution $N(2\sigma, \sigma^2)$, then σ is a scale parameter, and find a formula for L if you want a 97.5% one-sided confidence interval for σ.

(b) Show that if X has the distribution Unif$(0, \theta)$, then θ is a scale parameter. Find formulas for L and U that can be used if you know only $X = \max(X_1, X_2)$ and want a 92% confidence interval.

$\frac{X}{\sigma} \sim N(2,1)$ does not depend on σ, so is a scale parameter.

$$P\left(\frac{X}{\sigma} < q_{.975}\right) = 0.975$$

$$P\left(\frac{X}{\sigma} > q_{.975}\right) = 0.025$$

$$P\left(\sigma < \frac{X}{q_{.975}}\right) = 0.025$$

$$P(\sigma < L) = 0.025$$

$$L = \frac{X}{q_{.975}}$$

■

$\frac{X}{\theta} \sim U(0,1)$ does not depend on θ, so is a scale parameter.

$$\frac{X}{\theta} \sim Beta(n, 1)$$

$$P\left(q_{.04} < \frac{X}{\theta} < q_{.96}\right) = 0.92$$

$$P\left(q_{.04} < \frac{X}{\theta}\right) = P\left(\frac{X}{\theta} < q_{.96}\right) = 0.96$$

$$P\left(q_{.04} > \frac{X}{\theta}\right) = P\left(\frac{X}{\theta} > q_{.96}\right) = 0.04$$

$$P\left(\sqrt{.04} > \frac{X}{\theta}\right) = P\left(\frac{X}{\theta} > \sqrt{.96}\right) = 0.04$$

Using quantiles from our beta distribution

$$P(\theta > 5X) = P\left(\theta < \frac{X}{\sqrt{.96}}\right) = 0.04$$

$$P(\theta > U) = P(\theta < L) = 0.04$$

$$U = 5X, L = \frac{X}{\sqrt{.96}}$$

■

35. 12-1

A modern-day Robinson Crusoe has been shipwrecked on an island which may lie off the coast of Brazil. Among the few items that he was able to salvage from his sinking ship are two books: one containing tables of the distribution functions for the standard normal distribution and for Student t distributions with various numbers of degrees of freedom and one named <u>Wild Plants of the Atlantic Islands</u>.

Like the original Crusoe, he makes raisins from the local wild grapes. From a random sample of n vines, he gets an average yield of $\overline{x}$ grams of raisins. From his botany book, he learns that on islands off the coast of Brazil, the yield of raisins has a normal distribution with mean μ and variance σ^2.

(a) How, given available resources, can he test the null hypothesis that his island is off the coast of Brazil?

(b) Suppose that the botany book includes only the value of μ, not of σ^2, but that modern Crusoe has calculated the sample standard deviation s of his raisin yields. How does he test the null hypothesis that his island is off the coast of Brazil?

Assume that the null hypothesis is true and use σ^2 as our variance

$$X \sim N(\mu, \sigma^2)$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Find the Z-score and look up its p-value from the book to determine if we should believe our null hypothesis to be true.

■

$$\frac{S^2}{\sigma^2} * (n-1) \sim \chi^2_{n-1}$$

$$T = \frac{\frac{\frac{\bar{X} - \mu}{\sigma}}{\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{(n-1)\sigma^2}}} = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$$

Find the t-statistic and look up its p-value from the book using $n - 1$ degrees of freedom to determine if we should believe our null hypothesis to be true.

■

36. 12-2

Likelihood ratio tests

A hypothesis (null or alternative) is *simple* if it completely specifies the distribution of the population. In this case, the Neyman-Pearson Lemma (not proved) asserts that the most powerful test is one in which the likelihood ratio for the competing hypotheses is set to some value c (perhaps 1).

(Example 8.18) You have been buying Ukrainian cyberpets, whose lifetime in years is an exponential random variable whose density function is $f(X;\lambda) = \lambda e^{-\lambda X}$ with $\lambda = 8$. You suspect that the pet server has been taken over by Russians, who use $\lambda = 10$.

(a) Design a maximum-likelihood test, assuming that you have observed the lifetimes of nine cyberpets.

(b) How would you arrange for the probabillity of a Type I error to be .05?

(c) How would you then calculate the probability of a Type II error?

(d) How would a Bayesian approach this problem?

$$L(X, \lambda = 8) = \prod_{i=1}^{9} 8e^{-8X_i}$$

$$L(X, \lambda = 10) = \prod_{i=1}^{9} 10e^{-10X_i}$$

$$\frac{\prod_{i=1}^{9} 8e^{-8X_i}}{\prod_{i=1}^{9} 10e^{-10X_i}} = c$$

c is a constant

$$\frac{8^9}{10^9} e^{2\sum_{i=1}^{9} X_i} = c$$

$$.8^9 e^{2\sum_{i=1}^{9} X_i} = c$$

Reject if:

$$.8^9 e^{2\sum_{i=1}^{9} X_i} < c$$

■

$$\sum_{i=1}^{9} X_i < \frac{1}{2}\ln\left(\frac{c}{.8^9}\right)$$

$$c_2 = \frac{1}{2}\ln\left(\frac{c}{.8^9}\right)$$

$$P\left(\sum_{i=1}^{9} X_i < c_2 | \lambda = 8\right) = 0.05$$

We then find the 0.05 quantile of the Gamma(9, 8) distribution for our c_2 since

$$\sum_{i=1}^{9} X_i \sim Gamma(9,8)$$

■

The probability of a Type II error would be:

$$P\left(\sum_{i=1}^{9} X_i > c_2 | \lambda = 10\right)$$

■

A Bayesian approach would be to start with $\lambda_{prior} = 8$ and then update λ based on the observed X_i's.

■