Friday's lecture

Joint densities
Marginal distributions
The distribution of a ratio
Problems

Problem solutions

1. $E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \right) dx$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\,dx\,dy$$

2. $E(X) = \sum_{k=1}^{\infty} k p_X(k) = p_X(1) + 2p_X(2) + \ldots$

$$= P(X>0) + P(X>1) + P(X>2) + \ldots$$

$$= \sum_{k=0}^{\infty} P(X>k)$$

3. $P(X<1000, Y<1000) = \int_0^{1000}\int_0^{1000} \lambda^2 e^{-\lambda(x+y)}\,dx\,dy$

$$= (1-\exp(-1000\lambda))^2$$

$$(1-e^{-1000\lambda})^2 = 0.01 \Rightarrow \lambda = 0.00011$$

4. From 2, if X and Y are non-negative integer-valued, we have

$$E(X) = \sum_{k=0}^{\infty}(1-F_X(k)) \geq \sum_{k=0}^{\infty}(1-F_Y(k)) = E(Y)$$

2 can be generalized for integer-valued random variables to

$$E(X) = \sum_{k=0}^{\infty} P(X>k) + \sum_{k=-\infty}^{-1} P(X \leq k)$$

and almost the same argument applies.

In the continuous case, if $X \geq 0$, we have

$$\int_0^{\infty}(1-F_X(x))\,dx = \int_0^{\infty}\int_x^{\infty} f_X(t)\,dt\,dx$$

$$= \int_0^{\infty} f_X(t)\int_0^{t} dx = \int_0^{\infty} t f_X(t)\,dt = E(X)$$

and the same argument as in the first case applies. Finally, the second argument can similarly be extended to the continuous case.

5. (a) $P(S=i \mid N=k) = \binom{k}{i} \pi^i (1-\pi)^{k-i}$

(b) $P(S=i, N=k) = P(S=i \mid N=k)P(N=k)$

$$= \binom{k}{i} \pi^k (1-\pi)^{i-k} \binom{n}{k} p^k (1-p)^{n-k}$$

(c)

$$P(S=j)=\sum_{i=j}^{n}P(S=j,N=i)=\sum_{i=j}^{n}\frac{n!\,\pi^{j}(1-\pi)^{i-j}p^{i}(1-p)^{n-i}}{(n-i)!\,j!(i-j)!}$$

$$=(1-p)^{n}\left(\frac{\pi}{1-\pi}\right)^{j}\binom{n}{j}\sum_{i=j}^{n}\binom{n-j}{j-i}\left(\frac{(1-\pi)p}{1-p}\right)^{i}$$

$$=(1-p)^{n}\left(\frac{\pi}{1-\pi}\right)^{j}\binom{n}{j}\left(\frac{(1-\pi)p}{1-p}\right)^{j}\left(1+\frac{(1-\pi)p}{1-p}\right)^{n-j}$$

$$=\binom{n}{j}(\pi p)^{j}(1-\pi p)^{n-j}$$

(d)

$$P(N=i|S=j)=\frac{P(N=i,S=j)}{P(S=j)}$$

$$=\binom{n-j}{n-i}\left(\frac{1-p}{1-\pi p}\right)^{n-i}\left(1-\frac{1-p}{1-\pi p}\right)^{i-j}$$

156

A conditional density

If (X,Y) has joint density $f_{X,Y}(x,y)$, we can define a conditional density of X, given that Y=y, by

$$f_{X|Y}(x\,|\,y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}$$

We can then compute

$$P(X\in A\,|\,Y=y)=\int_{x\in A}f_{X|Y}(x\,|\,y)\,dx$$

even though the condition {Y=y} has probability 0.
Discrete case?

157

An example

Let $f_{X,Y}(x,y)=2$, $x\geq 0$, $y\geq 0$, $x+y\leq 1$. Find the conditional density of Y given that X=x.

158

Independence

Two random variables are *independent* if

$$P(X\in A,Y\in B)=P(X\in A)P(Y\in B)$$

In particular, this holds if

$$p_{X,Y}(x,y)=p_X(x)p_Y(y)$$

or

$$f_{X,Y}(x,y)=f_X(x)f_Y(y)$$

or

$$f_{X|Y}(x|y)=f_X(x)$$

159

Independence

Two random variables are *independent* if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

In particular, this holds if

$$p_{X,Y}(x,y) = p_X(x)p_Y(y)$$

or

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

or

$$f_{X|Y}(x|y) = f_X(x)$$

160

The addition rule for expectations

$E(X+Y) = E(X) + E(Y)$

NOTE: No assumption of independence. This result holds whenever the expectations exist.

A special case: $E(aX + b) = a\,E(X) + b$

161

The addition rule for variances

$Var(X+Y) = E((X+Y)^2) - (E(X+Y))^2$
$\quad = E(X^2)+2E(XY)+E(Y^2) - (E(X))^2$
$\quad - 2E(X)E(Y) - (E(Y))^2$
$\quad = Var(X) + Var(Y)$
$\quad + 2(E(XY) - E(X)E(Y))$

If X and Y are independent,

$$E(XY) = \iint xyf_X(x)f_Y(y)\,dx\,dy$$

$$= \int xf_X(x)\,dx \int yf_Y(y)\,dy = E(X)E(Y)$$

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

162

Covariance

The covariance of X and Y is defined as

$Cov(X,Y) = E\{(X - E(X))(Y - E(Y))\}$
$\quad = E(XY) - E(X)E(Y) - E(X)E(Y)$
$\quad + E(X)E(Y)$
$\quad = E(XY) - E(X)E(Y)$

If X and Y are independent, $Cov(X,Y) = 0$.

$Cov(X,X) =$

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

$Var(X-Y) =$

163

Correlation

Units of covariance is product of units of X and Y. Sometimes one wants a unit-free quantity. To do that we *standardize* X and Y:

$$X^* = \frac{X - E(X)}{\sqrt{Var(X)}},\ Y^* = \frac{Y - E(Y)}{\sqrt{Var(Y)}}$$

Define the correlation coefficient

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

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

164

Properties of correlation coefficient

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

If $|\rho(X,Y)| = 1$ then Y = aX + b

165

Last Monday's lecture

Conditional distribution and density
Independent random variables
The addition rules for expected value and variance
Covariance and correlation

166

An example

Let X be -1, 0 , or 1 with equal probabilties 1/3.
E(X) =
Let Y = 1 if X=0, 0 otherwise.
E(Y) =
XY =
E(XY) =
Cov(X,Y) = E(XY) – E(X)E(Y) =

Are X and Y independent?

167

Calculating covariance and correlation

$$f_{X,Y}(x,y) = 2,\ 0 < x < y < 1$$
$$f_X(x) = 2(1-x),\ 0 < x < 1$$
$$f_Y(y) = 2y,\ 0 < y < 1$$
$$E(X) = \int_0^1 x\, 2(1-x)\, dx = \tfrac{1}{3}$$
$$E(Y) = \int_0^1 y\, 2y\, dy = \tfrac{2}{3}$$
$$E(XY) = \int_0^1 \int_0^y xy\, 2\, dx\, dy = \tfrac{1}{4}$$
$$\mathrm{Cov}(X,Y) = \tfrac{1}{4} - \tfrac{1}{3}\tfrac{2}{3} = \tfrac{1}{36}$$
$$\mathrm{Var}(X) = \int_0^1 x^2\, 2(1-x)\, dx - \left(\tfrac{1}{3}\right)^2 = \tfrac{1}{18}$$
$$\mathrm{Var}(Y) = \int_0^1 y^2\, 2y\, dy - \left(\tfrac{2}{3}\right)^2 = \tfrac{1}{18}$$
$$\mathrm{Corr}(X,Y) = \frac{1/36}{1/18} = \tfrac{1}{2}$$

168

Law of large numbers

Let X_1, X_2, ... be independent random variables, all with the same distribution having expected value μ and variance σ^2. Then

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| > \varepsilon\right) \to 0 \text{ as } n \to \infty$$
$$E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) =$$
$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) =$$

Link

We write $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, the *sample average.*

169

Estimation

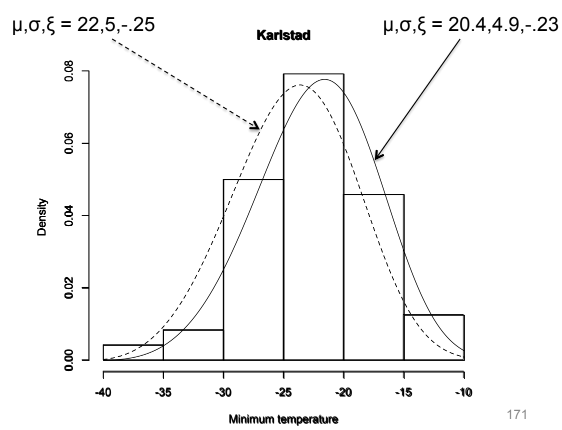
I have studied minimum annual temperatures in Karlstad, Sweden. It has been suggested that

$$F_X(-x;\mu,\sigma,\xi) = \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]\right\}$$

If we knew the *parameters* ξ, μ and σ, we could draw a histogram of the data and plot the corresponding density to see if it is a good fit.

170

Karlstad data



171

Monday's lecture

Covariance and correlation
Law of large numbers
Estimation

Concepts

Theoretical	*Practical*
Parameter θ	Parameter value θ_0
Sample $X_1, X_2, \ldots, X_n$	Data $x_1, x_2, \ldots, x_n$
Statistic $g(X_1, \ldots, X_n)$	Observed statistic $g(x_1, \ldots, x_n)$
Estimator $\hat{\theta}(X_1, \ldots, X_n)$	Estimate $\hat{\theta}(x_1, \ldots, x_n)$
pdf $f_X(x;\theta)$	Estimated pdf $f_X(x;\hat{\theta})$

Sampling distribution

Since $\hat{\theta}(X_1, \ldots, X_n)$ is a random variable, we can compute its *sampling distribution* cdf $P(\hat{\theta}(X_1, \ldots, X_n) \le x)$ and other properties such as

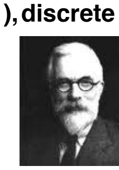
$$E_\theta(\hat{\theta}(X_1, \ldots, X_n)) =$$
$$\int \cdots \int \hat{\theta}(x_1, \ldots, x_n) f_X(x_1, \ldots, x_n; \theta) dx_1 \ldots dx_n$$
$$\text{bias}(\hat{\theta}, \theta) = E_\theta(\hat{\theta}) - \theta$$
$$\text{Var}_\theta(\hat{\theta})$$
$$\text{se}(\hat{\theta}) = \text{sd}_\theta(\hat{\theta}) = h(\theta)$$
$$\text{ese}(\hat{\theta}) = h(\hat{\theta})$$

What estimation is all about

In 1918 R. A. Fisher proposed estimating parameters by considering
how likely are the data if θ is the true parameter?
Choosing the parameter that makes the observations most likely is formalized using the *likelihood function*

$$L(\theta) = \begin{cases} f_{X_1, \ldots, X_n}(\boxed{x_1, \ldots, x_n}; \boxed{\theta}), & \text{continuous case} \\ p_{X_1, \ldots, X_n}(x_1, \ldots, x_n; \theta), & \text{discrete case} \end{cases}$$

The data are *fixed*
The parameter is *varying*

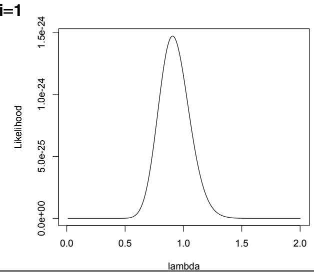


R.A.Fisher
1890-1962

The exponential case

Consider $x_1, \ldots, x_n$ independent observations from an exponential distribution with parameter λ.
The likelihood is

$$L(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda x_i) = \lambda^n \exp(-\lambda \sum_{i=1}^{n} x_i)$$



176

The method of maximum likelihood

Define the mle $\hat{\theta} = \arg\max(L(\theta))$
We compute it by setting $L'(\theta) = 0$ and checking that $L''(\theta) < 0$, or that L' has sign change + 0 – about the maximum. Alternatively, plot $L'(\theta)$ as a function of θ and find the maximum numerically.
Computational trick: maximize the *log likelihood* $\ell(\theta) = \log(L(\theta))$

177

Exponential case

$$L(\lambda) = \lambda^n e^{-\lambda \sum x_i}$$

$$L'(\lambda) = n\lambda^{n-1} e^{-\lambda \sum x_i} - \lambda^n \left(\sum x_i\right) e^{-\lambda \sum x_i}$$

$$= \lambda^{n-1} e^{-\lambda \sum x_i} (n - \lambda \sum x_i)$$

$$\ell(\lambda) = n \log(\lambda) - \lambda \sum x_i$$

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum x_i = 0 \Rightarrow \hat{\lambda} = 1/\bar{x}$$

$$\ell''(\lambda) = -\frac{n}{\lambda^2} < 0$$

178

Binomial case

$$L(p) = \prod_{i=1}^{n} \binom{N}{x_i} p^{x_i} (1-p)^{N-x_i} = \text{const} \times p^{\sum x_i} (1-p)^{Nn - \sum x_i}$$

$$\ell(p) = \text{const} + \log \frac{p}{1-p} \sum x_i + \log(1-p) Nn$$

$$\ell'(p) = \frac{\sum_{i=1}^{n} x_i}{p(1-p)} - \frac{Nn}{1-p} = 0 \Rightarrow \hat{p} = \frac{\bar{x}}{N}$$

$$\ell'(p) = \frac{nN}{p(1-p)} (\hat{p} - p) \text{ so } + 0 -$$

179

Friday's lecture

The method of maximum likelihood
Computational tools
Checking for a maximum
Problems

Problem solutions

1. **Note misprint in problem:**

$$E(X|Y=y)=\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,dx$$

(a) $E(E(X|Y))=\int\int x\frac{f_{X,Y}(x,y)}{f_Y(y)}f_Y(y)\,dy\,dx=$

$$=\iint x f_{X,Y}(x,y)\,dx\,dy=E(X)$$

(b) $Var(X|Y=y)=E(X^2|Y=y)-\left[E(X|Y=y)\right]^2$

As in (a), $E(E(X^2|Y=y))=E(X^2)$

Let $Z=E(X|Y=y)$, so

$Var(X)=E(Var(X\,|\,Y=y))+\left[E(Z)\right]^2$

But $E(X)=E(E(X|Y=y))=E(Z)$ so $\left[E(X)\right]^2=\left[E(Z)\right]^2$ whence

$$Var(X)=E(Var(X|Y=y))+E(Z^2)-\left[E(Z)\right]^2$$

$$=E(Var(X|Y=y)+Var(E(X|Y=y))$$

(c)

$$E(Y)=E(E(Y|X=k))=E(X\times 0.1)=10\times 0.1=1$$

2. Cov(U,V) = Cov(X+Y,Y+Z) = Cov(X,Y) + Cov(X,Z) + Cov(Y,Y) + Cov(Y,Z) = Var(Y) =144
Var(U) = Var(X) + Var(Y) + Cov(X,Y) = 169
Var(V) = Var(Y) + Var(Z) + Cov(Y,Z) = 180

$$Corr(U,V)=\frac{144}{\sqrt{169\times 180}}=0.83$$

3. (a) $E(\tilde{\mu})=E(\sum_{i=1}^{n}a_iX_i)=\sum_{i=1}^{n}a_iE(X_i)=\mu\sum_{i=1}^{n}a_i=\mu$

(b) $Var(\tilde{\mu})=\sum_{i=1}^{n}a_i^2Var(X_i)=\sigma^2\sum_{i=1}^{n}a_i^2$

(c) By symmetry the weights ought to be equal, in which case they each have to be 1/n. This is indeed optimal, since

$$\sum_{i=1}^{n}\left(a_i-\frac{1}{n}\right)^2=\sum_{i=1}^{n}a_i^2-\frac{2}{n}\sum_{i=1}^{n}a_i+\frac{n}{n^2}=\sum_{i=1}^{n}a_i^2-\frac{1}{n}$$

so $\sum_{i=1}^{n}a_i^2=\sum_{i=1}^{n}\left(a_i-\frac{1}{n}\right)^2+\frac{1}{n}\geq\frac{1}{n}$

with equality if and only if each a_i = 1/n.

4. Cov(X,aX+b) = a Cov(X,X) = a Var(X)
Var(aX+b) = a^2Var(x) so Corr(X,aX+b) = $a/|a|$. Conversely, if Corr(X,Y) = 1 write Var(Y* – X*) = Var(Y*) + Var(X*) – 2 = 0 so Y* – X* = c or

$$\frac{Y-E(Y)}{\sigma_Y} = \frac{X-E(X)}{\sigma_X} + c$$

so Y = aX + b where

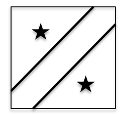
$$a = \frac{\sigma_Y}{\sigma_X}, b = E(Y) + \sigma_Y (c + \frac{E(X)}{\sigma_X})$$

The case Corr(X,Y)=-1 is simlar, starting from Var(Y* + X*) = 0

5. Let X and Y be the respective arrival times. They are independent, U(9,10). We need to compute P(|X–Y|>10) = 1– P(|X – Y| ≤ 10).

184

The joint distribution is uniform on the square with corners (9,9), (9,10), (10,9) and (10,10). The probability we want therefore is the area of the two triangles with ★ below:



This area is clearly $(5/6)^2 = 25/36$.

185

HOV lane needed?

The following data are for passenger car occupancy during one hour at Wilshire and Bundy in Los Angeles:

Occupants	Frequency	Predicted
1	676	
2	227	
3	56	
4	26	
5	6	
≥6	14	

The geometric distribution $p_X(y) = p(1-p)^{y-1}$ is a reasonable fit. The log likelihood is

$$l(p) = n \log(p) + \sum_{i=1}^{n} (y_i - 1)\log(1-p)$$

whence

$$l'(p) = \frac{n}{p} - \frac{\sum (y_i - 1)}{1-p} = 0 \quad \Rightarrow \quad \hat{p} = \frac{1}{\bar{y}}$$

186

HOV, cont.

But we do not have detailed data on the ≥6 group. However,

$$P(Y \geq 6) = \sum_{k=6}^{\infty} p(1-p)^{k-1} = p(1-p)^5 \sum_{k=0}^{\infty} (1-p)^k = (1-p)^5$$

so the log likelihood, the log probability of what we actually observed, becomes

$$\begin{aligned} l(p) &= \sum_{\{i: y_i < 6\}} (\log(p) + (y_i - 1)\log(1-p)) + \sum_{\{i: y_i \geq 6\}} 5\log(1-p) \\ &= (1011 - 14)\log(p) + 455\log(1-p) + 14 \times 5\log(1-p) \\ &= 997\log(p) + 525\log(1-p) \end{aligned}$$

187

The uniform distribution

Let $X_1,\ldots,X_n$ be iid $U(0,\theta)$. Then
$L(\theta) = \theta^{-n}$, so
$l(\theta) = -n \log(\theta)$ and
$l'(\theta) = -n/\theta$
What is the mle?

188

Reparametrization

A drug reaction surveillance program was carried out in 9 hospitals. Out of 11,526 monitored patients, 3,240 had an adverse reaction.
Model: X = # adverse reactions ~
Log likelihood

Standard error of the mle:
The *bootstrap method* just plugs in the estimate of p into the formula for the standard error:
But the standard error is a function of p. What is its mle?
Fact: The mle of $h(\theta)$ is $h(\hat{\theta})$.

189

Stopping

Flip a coin until the first heads. Suppose it takes 6 tries. The likelihood is

Now suppose we were going to flip the coin 6 times, and happened to get one head. The likelihood is

How are the mles different?

Fact: Changing the likelihood by a constant does not change the mle.

However, the standard errors would be different.

190