B555: Homework 2

1. Suppose that X is a discrete n-dimensional random vector

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

and that $a \in \Re^n$ (a is a constant n-dimensional vector). Show $E(a^t X) = a^t E(X)$ where a^t denotes the transpose of a. (The result also holds for continuous random vectors).

Solution

Consider first case where $X_1, \ldots, X_n$ are discrete random variables with joint distribution $p(x_1, \ldots, x_n)$.

$$\begin{aligned} E(a^t X) &= \sum_{x_1,\ldots,x_n} (a_1 x_1 + \ldots, a_n x_n) p(x_1, \ldots, x_n) \\ &= a_1 \sum_{x_1,\ldots,x_n} x_1 p(x_1, \ldots, x_n) + \ldots + a_n \sum_{x_1,\ldots,x_n} x_n p(x_1, \ldots, x_n) \\ &= a_1 E X_1 + \ldots + a_n E X_n \\ &= a^t E X \end{aligned}$$

For continuous variables the argument is the same with summation replaced by integration.

2. Let X, Y be independent continuous random variables with density $f(x, y)$, where $f(x, y) = g(x)h(y)$ for some functions g, h. Show that X and Y are independent.

Solution

Note that $f_X(x) = \sum_y f(x, y) = \sum_y g(x)h(y) = g(x) \sum_y h(y)$ and, similarly, $f_Y(y) = h(y) \sum_x g(x)$. Then

$$\begin{aligned} f_X(x) f_Y(y) &= g(x)h(y) \sum_{x'} g(x') \sum_{y'} h(y') \\ &= f(x, y) \end{aligned}$$

since $1 = \sum_{x,y} f(x, y) = \sum_x \sum_y g(x)h(y) = \sum_x g(x) \sum_y h(y)$

3. For this problem, recall that random variables X, Y are conditionally independent given Z if X and Y are independent under the conditional distribution $f(x, y|z)$ (the distribution of X, Y having observed $Z = z$.

Suppose that we consider random minutes from randomly chosen drivers to determine three binary random variables:

$$\begin{aligned} H &= \text{the driver is within 1 mile of home} \\ M &= \text{the driver is on a main road} \\ A &= \text{the driver has had an accident in the sampled minute} \end{aligned}$$

Suppose that $P(H = 1) = .8$ along with $P(M|H)$ and $P(A = 1|, H, M)$ in the following two tables.

$P(M\|H)$	$M = 0$	$M = 1$
$H = 0$	.1	.9
$H = 1$	.6	.4

$P(A = 1\|H, M)$	$M = 0$	$M = 1$
$H = 0$	.0003	.0008
$H = 1$	.0003	.0008

(a) Compute $P(A=1|H=0)$ and $P(A=1|H=1)$ Are accidents more likely near home?
Solution

$$\begin{aligned} P(A=1|H=0) &= \sum_m P(A=1, M=m|H=0) \\ &= \sum_m P(M=m|H=0)P(A=1|H=0, M=m) \\ &= (.1)(.0003) + (.9)(.0008) = .00075 \end{aligned}$$

$$\begin{aligned} P(A=1|H=1) &= \sum_m P(A=1, M=m|H=1) \\ &= \sum_m P(M=m|H=1)P(A=1|H=1, M=m) \\ &= (.6)(.0003) + (.4)(.0008) = .0005 \end{aligned}$$

Accidents are less likely near home.

(b) Compute $P(A=1|M=0)$ and $P(A=1|M=1)$. What conditional independence relation can you conclude?
Solution
Note that $p(a|m) = \sum_h p(a|h,m)p(h|m)$. In the sum the first factor doesn't depend on h as can be seen from the table, while the second factor sums to 1. So $p(a|m) = p(a|m,h)$ and A is conditionally independent of H given M. To see this more formally note that once we have $p(a|m) = p(a|m,h)$ then $p(a,h|m) = p(h|m)p(a|h,m) = p(h|m)p(a|m)$ which is the definition of H and A being conditionally indpendent given M

(c) Give a cause and effect hypothesis that is consistent with your conditional indepedence statement.
Solution
Accidents are caused, in part, by the kind of road one uses, with main roads being less dangerous and non-main roads being more dangerous. Once we know which type of road we are on, whether or not we are close to home is not relevant.

(d) Suppose that a driver is driving on a stretch of highway (main road) on her way home. As she crosses the 1-mile-to-home road marker does she encounter an elevated probability of accident?
Solution
Due to the conditional independence relationship there is no change in the probability of an accident.

4. Simulation of Binomial

(a) A random variable has $X \sim \text{Unif}(0,1)$ when its density function is

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

The R command "runif(n)" creates a vector of n independent $\text{Unif}(0,1)$ outcomes. Use this function to simulate a sequence of 10,000 independent $\text{Binomial}(500, 1/2)$ random variables and plot the resulting histogram. You can plot a histogram of a data vector x with "hist(x)".

(b) Letting your sequence of Binomial variables be $X_1, \ldots, X_n$, "standardize" these variables by taking $Y_n = (X_n - \mu)/\sigma$ where μ and σ are the mean and standard deviation of the variables. What are the mean and variance of the Y_n variables?

(c) Now simulate 10,000 *standardized* Binomials and plot the empirical cumulative distribution function $\hat{F}(x)$ where

$$\hat{F}(x) = \hat{P}(X \leq x) = \frac{|\{i : X_i \leq x\}|}{n}$$

Use "plot(x,type='l')" to plot with a solid line

(d) On the same plot show the cumulative distribution function of the standard normal. "pnorm(x)" gives the probability that a $N(0,1)$ variable is less than x. You can add to an existing plot with "lines(x)." Your two plots should look very similar. If X is a Binomial with large n, the fact that

$$\frac{X-np}{\sqrt{np(1-p)}} \overset{\text{approx}}{\sim} N(0,1)$$

is known as the normal approximation to the Binomial.

Solution

Posted on Canvas as sim_binom.r

5. It is well-known that independent random variables are also uncorrelated.

Consider the variables $X=(X_1,X_2)$ having probability density function

$$p(x) = \begin{cases} \frac{1}{\pi} & x_1^2+x_2^2<1 \\ 0 & \text{otherwise} \end{cases}$$

Show that X_1 and X_2 are uncorrelated but not independent.

Solution

The marginal distribution of both X_1 and X_2 are "supported (have non-0 density) on the interval $(-1,1)$. Thus the product of the marginals will be supported on $(-1,1)x(-1,1)$. Since the joint density is supported on a subset of $(-1,1)x(-1,1)$ the joint density cannot be the product of the marginal densities, so not independent. To see uncorrelated:

$$\begin{aligned}
\text{Cov}(X_1,X_2) &= \int x_1,x_2 p(x_1,x_2)dx_1dx_2 \\
&= \int_{-1}^{1}\int_{-\sqrt{1-x_1^2}}^{\sqrt{1-x_1^2}} \frac{1}{\pi}x_1x_2dx_2dx_1 \\
&= \int_{-1}^{1} x_1\frac{2\sqrt{1-x_1^2}}{\pi} \\
&= 0
\end{aligned}$$

6. Suppose that $X=(X_1,X_2)\sim N(\mu,\Sigma)$ where

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

(a) Show that in this case X_1,X_2 uncorrelated implies they are also independent. It might help to know that the marginal distribution of X_1 is normal with mean μ_1 and variance $\sigma_1^2=\sigma_{11}$ with a simliar statement for X_2.

Solution

If X_1,X_2 uncorrelated then $\sigma_{12}=\sigma_{21}=0$. Thus

$$\begin{aligned}
p(x_1,x_2) &= \frac{1}{(2\pi)|\Sigma|^{1/2}}e^{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)} \\
&= ce^{-\frac{1}{2}(x_1-\mu_1)\sigma_{11}^{-1}(x_1-\mu_1)}e^{-\frac{1}{2}(x_2-\mu_2)\sigma_{22}^{-1}(x_2-\mu_2)} \\
&= = g(x_1)h(x_2)
\end{aligned}$$

We have seen that if the joint pdf can be expressed as a product of functions that depend only on the individual variables, the variables must be independent.

(b) Suppose, $X = (X_1, X_2, \ldots, X_p) \sim N(\mu, \Sigma)$. Show that if $\Sigma^{-1}_{ij} = 0$ then X_i and X_j are conditionally independent given *all* the other variables. (One can also show the stated conditional independence assumption implies $\Sigma^{-1}_{ij} = 0$ for normal random variables).
Solution
It may help to first consider a 3-variable example. For notation write $S = \Sigma^{-1}$ and suppose that $S_{12} = 0$. Then $p(x_1, x_2|x_3)$ proportional to the joint pdf, $p(x_1, x_2, x_3)$ when x_3 is fixed. Thus

$$\begin{aligned}
p(x_1, x_2|x_3) &= ce^{-\frac{1}{2}\sum_{i,j=1}^{3} S_{ij}(x_i-\mu)(x_j-\mu_j)} \\
&= ce^{-\frac{1}{2}(S_{11}(x_1-\mu_1)^2+S_{22}(x_2-\mu_2)^2+S_{33}(x_3-\mu_3)^2+2S_{13}(x_1-\mu_1)(x_3-\mu_3)+2S_{23}(x_2-\mu_2)(x_3-\mu_3))} \\
&= c'e^{-\frac{1}{2}(S_{11}(x_1-\mu_1)^2+2S_{13}(x_1-\mu_1)(x_3-\mu_3))}e^{-\frac{1}{2}(S_{22}(x_2-\mu_2)^2+2S_{13}(x_2-\mu_2)(x_3-\mu_3))} \\
&= c'g(x_1)h(x_2)
\end{aligned}$$

where the S_{33} terms has been absorbed into c'. In the general case suppose $S_{i'j'} = 0$. Then

$$\begin{aligned}
p(x_{i'}, x_{j'}|x_k : k \neq i', j') &= ce^{-\frac{1}{2}\sum_{i,j} S_{ij}(x_i-\mu)(x_j-\mu_j)} \\
&= ce^{-\frac{1}{2}\sum_{i\neq i', j\neq j'} S_{ij}(x_i-\mu_i)(x_j-\mu_j)} \\
&\quad \times e^{-\frac{1}{2}\sum_{i=i', j\neq j'} S_{ij}(x_i-\mu_i)(x_j-\mu_j)} \\
&\quad \times e^{-\frac{1}{2}\sum_{i\neq i', j=j'} S_{ij}(x_i-\mu_i)(x_j-\mu_j)}
\end{aligned}$$

The x terms in the first factor are constant hence this term is constant. The terms in the 2nd factor involve $x_{i'}$ and other constant terms, hence this factor is a function of $x_{i'}$ only. Similarly, the 3rd factor is a function of $x_{j'}$ only. Thus we have shown that

$$p(x_{i'}, x_{j'}|x_k : k \neq i', j') = g(x_{i'})h(x_{j'})$$

.

(c) Suppose $X_1, X_2, \ldots, X_n$ defined by $X_1 = Z_1$ and $X_i = X_{i-1} + Z_i$ for $i = 2, \ldots, n$ where $Z_1, \ldots, Z_n \overset{\text{iid}}{\sim} N(0, 1)$. This is known as a random walk.

i. Show that $(X_1, \ldots, X_n) \sim N(0, \Sigma)$ for some covariance matrix Σ.
Solution
We have

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = A \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$$

where A is the lower-triangular matrix of 1's (with 1's also on the diagonal). Thus the vector X is joint normal.

ii. Without computing argue which elements of Σ^{-1} are 0.
Solution
By construction we have X_i and X_j are conditionally independent given all other variables, as long as $|i - j| > 1$. Thus $S_{i,j} =$ as long as $|i - j| > 1$.

7. Suppose that $X \sim \text{Unif}(0, 1)$.

(a) Show that the cdf of $Y = \frac{-\log(X)}{\lambda} = 1 - e^{-\lambda y}$ and conclude that $Y \sim \text{Exponential}(\lambda)$.
Solution

$$\begin{aligned}
P(Y \leq y) &= P(\frac{-\log(X)}{\lambda} \leq y) \\
&= P(X \geq e^{-\lambda y}) \\
&= 1 - e^{-\lambda y}
\end{aligned}$$

which is the cdf for the Exponential distribution with parameter λ.

(b) Suppose that $X_1, X_2 \ldots \overset{\text{iid}}{\sim}$ Exponential(λ). Interpret X_1 as the time we need to wait for the first event, X_2 as the *additional* time we wait for the 2nd event, X_3 as the additional time for the 3rd event and so on. In R take $\lambda = 3$ and simulate 100,000 times the number of events that occur in the interval $[0, 1]$. Compute the empirical pmf, $\hat{p}(n)$ for $n = 0, 1, \ldots$ by

$$\hat{p}(n) = \frac{\#\text{ times experiments gives } n}{100000}$$

On the same plot compare these probabilities with the Poisson(λ) distribution. The process $N(t)$ that counts the number of events that occur before time t is known as a Poisson process.

Solution

This is on Canvas as poisson_process.r

8. Let M be a discrete variable with probability mass function $p(m)$. Suppose X is a continuous random vector that *depends* on M. Specifically, assume $X|M = m$ is $N(\mu_m, \Sigma_m)$.

(a) Even though M is discrete and X is continuous we can still get the joint distribution of M, X as $p(m, x) = p(m)p(x|m)$. In this problem we'll use the term density for dicrete, continuous and mixed cases. Compute

i. The joint density $p(m, x)$

Solution

$$p(m, x) = p(m)N(x; \mu_m, \Sigma_m)$$

where $N(x; \mu, \Sigma)$ is the multivariate normal density function with mean μ and covariance Σ.

ii. The marginal density $p(x)$.

Solution

$$p(x) = \sum_m p(m)N(x; \mu_m, \Sigma_m)$$

iii. The conditional density $p(m|x)$

Solution

$$p(m|x) = \frac{p(m)N(x; \mu_m, \Sigma_m)}{\sum_{m'} p(m')N(x; \mu_{m'}, \Sigma_{m'})}$$

(b) Suppose that a population is composed of C classes $1, \ldots, C$. let $p(i)$ be the proportion of the population in class i so $p(1) + \ldots + p(C) = 1$. Suppose we take several measurments, X, on the population and that these measurements behave differently under the classes. We model the "class conditional" distribution of X as multivariate normal with mean μ_i and covariance Σ_i.

i. If we observe $X = x$ give the conditional probabilities $P(M = m|X = x)$.

Solution

Given in the previous problem

ii. Suppose that you observe $X = x$ and want to choose the value of m, $\hat{M}$, that has the best chance of being right. What should $\hat{M}$ be? **Solution**

$$\hat{M} = \arg\max_m p(m)N(x; \mu_m, \Sigma_m)$$