STAT 801 Lecture 3

Reading for Today's Lecture:
Goals of Today's Lecture:

- Learn properties of independence
- Define Multivariate Normal Dist.
- Do 1 sample normal distribution theory.

Review of end of last time

I defined independent events and then independent random variables:
Definition: Events A_i, $i = 1, \ldots, p$ are **independent** if

$$P(A_{i_1} \cdots A_{i_r}) = \prod_{j=1}^{r} P(A_{i_j})$$

for any set of distinct indices $i_1, \ldots, i_r$ between 1 and p.
Definition: Random variables $X_1, \ldots, X_p$ are **independent** if

$$P(X_1 \in A_1, \cdots, X_p \in A_p) = \prod P(X_i \in A_i)$$

for any choice of $A_1, \ldots, A_p$.

Theorem 1 *1. If X and Y are independent then*

$$F_{X,Y}(x, y) = F_X(x) F_Y(y)$$

for all x, y

2. If X and Y are independent and have joint density $f_{X,Y}(x, y)$ then X and Y have densities, say f_X and f_Y, and

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

3. If X and Y are independent and have marginal densities f_X and f_Y then (X, Y) has joint density $f_{X,Y}(x, y)$ given by

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

4. *If*
$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$
for **all** x, y *then* X *and* Y *are independent.*

5. *If* (X, Y) *has density* $f(x, y)$ *and there are functions* $g(x)$ *and* $h(y)$ *such that*
$$f(x,y) = g(x)h(y)$$
for **all** *(well technically almost all)* (x, y) *then* X *and* Y *are independent and they each have a density given by*
$$f_X(x) = g(x)/\int_{-\infty}^{\infty} g(u)du$$
and
$$f_Y(y) = h(y)/\int_{-\infty}^{\infty} h(u)du\,.$$

New material
Proof:

1. Since X and Y are independent so are the events $X \le x$ and $Y \le y$; hence
$$P(X \le x, Y \le y) = P(X \le x)P(Y \le y)$$

2. For clarity suppose X and Y are real valued. In assignment 2 I have asked you to prove that the existence of $f_{X,Y}$ implies that f_X and f_Y exist (and are given by the marginal density formula). Then for any sets A and B
$$\begin{aligned} P(X \in A, Y \in B) &= \int_A \int_B f_{X,Y}(x,y)dydx \\ P(X \in A)P(Y \in B) &= \int_A f_X(x)dx \int_B f_Y(y)dy \\ &= \int_A \int_B f_X(x)f_Y(y)dydx \end{aligned}$$
Since $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ we see that for any set A and B
$$\int_A \int_B [f_{X,Y}(x,y) - f_X(x)f_Y(y)]dydx = 0$$
It follows (via measure theory) that the quantity in [] is 0 (for almost every pair (x, y)).

3. For any A and B we have

$$\begin{aligned} P(X \in A, Y \in B) &= P(X \in A)P(Y \in B) \\ &= \int_A f_X(x)dx \int_B f_Y(y)dy \\ &= \int_A \int_B f_X(x)f_Y(y)]dydx \end{aligned}$$

   If we **define** $g(x,y) = f_X(x)f_Y(y)$ then we have proved that for $C = A \times B$

$$P((X,Y) \in C) = \int_C g(x,y)dydx$$

   To prove that g is the joint density of (X,Y) we need only prove that this integral formula is valid for an arbitrary Borel set C, not just a rectangle $A \times B$. This is proved via a *monotone class* argument. You prove that the collection of sets C for which the identity holds has closure properties which guarantee that this collection includes the Borel sets.

4. This is proved via another monotone class argument.

5.

$$\begin{aligned} P(X \in A, Y \in B) &= \int_A \int_B g(x)h(y)dydx \\ &= \int_A g(x)dx \int_B h(y)dy \end{aligned}$$

   Take $B = R^1$ to see that

$$P(X \in A) = c_1 \int_A g(x)dx$$

   where $c_1 = \int h(y)dy$. From the definition of density we see that $c_1 g$ is the density of X. Since $\int\int f_{X,Y}(xy)dxdy = 1$ we see that $\int g(x)dx \int h(y)dy = 1$ so that $c_1 = 1/\int g(x)dx$. A similar argument works for Y.

Theorem 2 *If $X_1, \ldots, X_p$ are independent and $Y_i = g_i(X_i)$ then $Y_1, \ldots, Y_p$ are independent. Moreover, $(X_1, \ldots, X_q)$ and $(X_{q+1}, \ldots, X_p)$ are independent.*

Conditional probability

Def'n: $P(A|B) = P(AB)/P(B)$ provided $P(B) \neq 0$.

Def'n: For discrete random variables X and Y the conditional probability mass function of Y given X is

$$\begin{aligned} f_{Y|X}(y|x) &= P(Y = y|X = x) \\ &= f_{X,Y}(x,y)/f_X(x) \\ &= f_{X,Y}(x,y)/\sum_t f_{X,Y}(x,t) \end{aligned}$$

For absolutely continuous X the problem is that $P(X = x) = 0$ for all x so how can we define $P(A|X = x)$ or $f_{Y|X}(y|x)$? The solution is to take a limit

$$P(A|X = x) = \lim_{\delta x \to 0} P(A|x \le X \le x + \delta x)$$

If, for instance, X, Y have joint density $f_{X,Y}$ then with $A = \{Y \le y\}$ we have

$$P(A|x \le X \le x + \delta x)\&\frac{P(A \cap x \le X \le x + \delta x)}{P(x \le X \le x + \delta x)}$$

$$= \frac{\int_{-\infty}^{y} \int_{x}^{x+\delta x} f_{X,Y}(u,v)\,du\,dv}{\int_{x}^{x+\delta x} f_X(u)\,du}$$

Divide the top and bottom by δx and let δx tend to 0. The denominator converges to $f_X(x)$ while the numerator converges to

$$\int_{-\infty}^{y} f_{X,Y}(x,v)\,dv$$

So we define the conditional cdf of Y given $X = x$ to be

$$P(Y \le y|X = x) = \frac{\int_{-\infty}^{y} f_{X,Y}(x,v)\,dv}{f_X(x)}$$

Differentiate with respect to y to get the definition of the conditional density of Y given $X = x$ namely

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

or in words "conditional = joint/marginal".

The Multivariate Normal Distribution

Def'n: $Z \in R^1 \sim N(0,1)$ iff

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Def'n: $Z \in R^p \sim MVN(0,I)$ iff $Z = (Z_1, \ldots, Z_p)^t$ (a column vector for later use) with the Z_i independent and each $Z_i \sim N(0,1)$.

In this case according to our theorem

$$\begin{aligned} f_Z(z_1, \ldots, z_p) &= \prod \frac{1}{\sqrt{2\pi}} e^{-z_i^2/2} \\ &= (2\pi)^{-p/2} \exp\{-z^t z/2\} \end{aligned}$$

where the superscript t denotes matrix transpose.

Def'n: $X \in R^p$ has a multivariate normal distribution if it has the same distribution as $AZ + \mu$ for some $\mu \in R^p$, some $p \times p$ matrix of constants A and $Z \sim MVN(0,I)$.

If the matrix A is singular then X will not have a density. If A is invertible then we can derive the multivariate normal density by the change of variables formula:

$$X = AZ + \mu \Leftrightarrow Z = A^{-1}(X - \mu)$$

$$\frac{\partial X}{\partial Z} = A \qquad \frac{\partial Z}{\partial X} = A^{-1}$$

So

$$\begin{aligned} f_X(x) &= f_Z(A^{-1}(x-\mu)) |\det(A^{-1})| \\ &= \frac{(2\pi)^{-p/2} \exp\{-(x-\mu)^t (A^{-1})^t A^{-1}(x-\mu)/2\}}{|\det A|} \end{aligned}$$

Now define $\Sigma = AA^t$ and notice that

$$\Sigma^{-1} = (A^t)^{-1} A^{-1} = (A^{-1})^t A^{-1}$$

and

$$\det \Sigma = \det A \det A^t = (\det A)^2$$

Thus

$$f_X(x) = (2\pi)^{-p/2} (\det \Sigma)^{-1/2} \exp\{-(x-\mu)^t \sigma^{-1} (x-\mu)/2\}$$

which is called the $MVN(\mu, \Sigma)$ density. Notice that the density is the same for all A such that $AA^t = \Sigma$. This justifies the notation $MVN(\mu, \Sigma)$.

For which vectors μ and matrices Σ is this a density? Any μ will work but if $x \in R^p$ then

$$\begin{aligned} x^t \Sigma x &= x^t A A^t x \\ &= (A^t x)^t (A^t x) \\ &= \sum_1^p y_i^2 \\ &\ge 0 \end{aligned}$$

where $y = A^t x$. The inequality is strict except for $y = 0$ which is equivalent to $x = 0$. Thus Σ is a positive definite symmetric matrix. Conversely, if Σ is a positive definite symmetric matrix then there is a square invertible matrix A such that $AA^t = \Sigma$ so that there is a $MVN(\mu, \Sigma)$ distribution. (The matrix A can be found via the Cholesky decomposition, for example.)

More generally we say that X has a multivariate normal distribution if it has the same distribution as $AZ + \mu$ where now we remove the restriction that A be non-singular. When A is singular X will not have a density because there will exist an a such that $P(a^t X = a^t \mu) = 1$, that is, so that X is confined to a hyperplane. It is still true, however, that the distribution of X depends only on the matrix $\Sigma = AA^t$ in the sense that if $AA^t = BB^t$ then $AZ + \mu$ and $BZ + \mu$ have the same distribution.

Properties of the MVN distribution

1: All margins are multivariate normal: if

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

then $X \sim MVN(\mu, \Sigma)$ implies that $X_1 \sim MVN(\mu_1, \Sigma_{11})$.

2: All conditionals are normal: the conditional distribution of X_1 given $X_2 = x_2$ is $MVN(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$

3: $MX+\nu \sim MVN(M\mu+\nu, M\Sigma M^t)$. That is, an affine transformation of a Multivariate Normal is normal.

Normal samples: Distribution Theory

Theorem 3 *Suppose $X_1,\ldots,X_n$ are independent $N(\mu,\sigma^2)$ random variables. (That is each satisfies my definition above in 1 dimension.) Then*

1. *The sample mean $\bar X$ and the sample variance s^2 are independent.*
2. *$n^{1/2}(\bar X-\mu)/\sigma \sim N(0,1)$*
3. *$(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$*
4. *$n^{1/2}(\bar X-\mu)/s \sim t_{n-1}$*

Proof: Let $Z_i=(X_i-\mu)/\sigma$. Then $Z_1,\ldots,Z_p$ are independent $N(0,1)$ so $Z=(Z_1,\ldots,Z_p)^t$ is multivariate standard normal. Note that $\bar X=\sigma\bar Z+\mu$ and $s^2=\sum(X_i-\bar X)^2/(n-1)=\sigma^2\sum(Z_i-\bar Z)^2/(n-1)$ Thus

$$\frac{n^{1/2}(\bar X-\mu)}{\sigma}=n^{1/2}\bar Z$$

$$\frac{(n-1)s^2}{\sigma^2}=\sum(Z_i-\bar Z)^2$$

and

$$T=\frac{n^{1/2}(\bar X-\mu)}{s}=\frac{n^{1/2}\bar Z}{s_Z}$$

where $(n-1)s_Z^2=\sum(Z_i-\bar Z)^2$.

Thus we need only give a proof in the special case $\mu=0$ and $\sigma=1$.

Step 1: Define

$$Y=(\sqrt{n}\bar Z, Z_1-\bar Z,\ldots,Z_{n-1}-\bar Z)^t\,.$$

(This choice means Y has the same dimension as Z.) Now

$$Y=\begin{bmatrix}\frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}}\\ 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n}\\ -\frac{1}{n} & 1-\frac{1}{n} & \cdots & -\frac{1}{n}\\ \vdots & \vdots & \vdots & \vdots\end{bmatrix}\begin{bmatrix}Z_1\\ Z_2\\ \vdots\\ Z_n\end{bmatrix}$$

or letting M denote the matrix

$$Y = MZ$$

It follows that $Y \sim MVN(0, MM^t)$ so we need to compute MM^t:

$$MM^t = \left[\begin{array}{c|cccc} 1 & 0 & 0 & \cdots & 0 \\ \hline 0 & 1-\frac{1}{n} & -\frac{1}{n} & \cdots & -\frac{1}{n} \\ \vdots & -\frac{1}{n} & \ddots & \cdots & -\frac{1}{n} \\ 0 & \vdots & \cdots & & 1-\frac{1}{n} \end{array}\right]$$
$$= \left[\begin{array}{c|c} 1 & 0 \\ \hline & \\ 0 & Q \end{array}\right]$$

Now it is easy to solve for Z from Y because $Z_i = n^{-1/2}Y_1 + Y_{i+1}$ for $1 \le i \le n-1$. To get Z_n we use the identity

$$\sum_{i=1}^{n}(Z_i - \bar{Z}) = 0$$

to see that $Z_n = -\sum_{i=2}^n Y_i + n^{-1/2}Y_1$. This proves that M is invertible and we find

$$\Sigma^{-1} \equiv (MM^t)^{-1} = \left[\begin{array}{c|c} 1 & 0 \\ \hline & \\ 0 & Q^{-1} \end{array}\right]$$

Now we use the change of variables formula to compute the density of Y. It is helpful to let $\mathbf{y}_2$ denote the $n-1$ vector whose entries are $y_2, \ldots, y_n$. Note that

$$y^t\Sigma^{-1}y = y_1^2 + \mathbf{y}_2^t Q^{-1}\mathbf{y_2}$$

Then

$$\begin{aligned} f_Y(y) =& (2\pi)^{-n/2}\exp[-y^t\Sigma^{-1}y/2]/|\det M| \\ =& \frac{1}{\sqrt{2\pi}}e^{-y_1^2/2} \times \frac{(2\pi)^{-(n-1)/2}\exp[-\mathbf{y}_2^t Q^{-1}\mathbf{y_2}/2]}{|\det M|} \end{aligned}$$

Notice that this is a factorization into a function of y_1 times a function of $y_2, \ldots, y_n$. Thus $\sqrt{n}\bar{Z}$ is independent of $Z_1 - \bar{Z}, \ldots, Z_{n-1} - \bar{Z}$. Since s_Z^2 is a function of $Z_1 - \bar{Z}, \ldots, Z_{n-1} - \bar{Z}$ we see that $\sqrt{n}\bar{Z}$ and s_Z^2 are independent.

Furthermore the density of Y_1 is a multiple of the function of y_1 in the factorization above. But the factor in question is the standard normal density so $\sqrt{n}\bar{Z}N(0,1)$.

We have now done the first 2 parts of the theorem. The third part is a homework exercise but I will outline the derivation of the χ^2 density.

Suppose that $Z_1,\ldots,Z_n$ are independent $N(0,1)$. We define the χ^2_n distribution to be that of $U = Z_1^2+\cdots+Z_n^2$. Define angles $\theta_1,\ldots,\theta_{n-1}$ by

$$\begin{aligned}
Z_1 &= U^{1/2}\cos\theta_1 \\
Z_2 &= U^{1/2}\sin\theta_1\cos\theta_2 \\
\vdots &= \vdots \\
Z_{n-1} &= U^{1/2}\sin\theta_1\cdots\sin\theta_{n-2}\cos\theta_{n-1} \\
Z_n &= U^{1/2}\sin\theta_1\cdots\sin\theta_{n-1}
\end{aligned}$$

(These are spherical co-ordinates in n dimensions. The θ values run from 0 to π except for the last θ whose values run from 0 to 2π.) Then note the following derivative formulas

$$\frac{\partial Z_i}{\partial U} = \frac{1}{2U}Z_i$$

and

$$\frac{\partial Z_i}{\partial \theta_j} = \begin{cases} 0 & j>i \\ -Z_i\tan\theta_i & j=i \\ Z_i\cot\theta_j & j<i \end{cases}$$

I now fix the case $n=3$ to clarify the formulas. The matrix of partial derivatives is

$$\left[\begin{array}{ccc}
U^{-1/2}\cos\theta_1/2 & -U^{1/2}\sin\theta_1 & 0 \\
U^{-1/2}\sin\theta_1\cos\theta_2/2 & U^{1/2}\cos\theta_1\cos\theta_2 & -U^{1/2}\sin\theta_1\sin\theta_2 \\
U^{-1/2}\sin\theta_1\sin\theta_2/2 & U^{1/2} cos\theta_1\sin\theta_2 & U^{1/2}\sin\theta_1\cos\theta_2
\end{array}\right]$$

The determinant of this matrix may be found by adding $2U^{1/2}\cos\theta_j/\sin\theta_j$ times column 1 to column $j+1$ (which doesn't change the determinant). The resulting matrix is lower triangular with diagonal entries (after a small amount of algebra) $U^{-1/2}\cos\theta_1/2$, $U^{1/2}\cos\theta_2/\cos\theta_1$ and $U^{1/2}\sin\theta_1/\cos\theta_2$. We multiply these together to get

$$U^{1/2}\sin(\theta_1)/2$$

which is non-negative for all U and θ_1. For general n we see that every term in the first column contains a factor $U^{-1/2}/2$ while every other entry has a factor $U^{1/2}$. Multiplying a column in a matrix by c multiplies the determinant by c so the Jacobian of the transformation is $u^{(n-1)/2}u^{-1/2}/2$ times some function, say h, which depends only on the angles. Thus the joint density of $U, \theta_1, \ldots \theta_{n-1}$ is

$$(2\pi)^{-n/2}\exp(-u/2)u^{(n-2)/2}h(\theta_1,\cdots,\theta_{n-1})/2$$

To compute the density of U we must do an $n-1$ dimensional multiple integral $d\theta_{n-1}\cdots d\theta_1$. We see that the answer has the form

$$cu^{(n-2)/2}\exp(-u/2)$$

for some c which we can evaluate by making

$$\int f_U(u)du = c\int u^{(n-2)/2}\exp(-u/2)du = 1$$

Substitute $y = u/2$, $du = 2dy$ to see that

$$c2^{(n-2)/2}2\int y^{(n-2)/2}e^{-y}dy = c2^{(n-1)/2}\Gamma(n/2) = 1$$

so that the χ^2 density is

$$\frac{1}{2\Gamma(n/2)}\left(\frac{u}{2}\right)^{(n-2)/2}e^{-u/2}$$

Finally the fourth part of the theorem is a consequence of the first 3 parts of the theorem and the definition of the t_ν distribution, namely, that $T \sim t_\nu$ if it has the same distribution as

$$Z/\sqrt{U/\nu}$$

where $Z \sim N(0,1)$, $U \sim \chi^2_\nu$ and Z and U are independent.

However, I now derive the density of T in this definition:

$$\begin{aligned} P(T \le t) &= P(Z \le t\sqrt{U/\nu}) \\ &= \int_0^\infty \int_{-\infty}^{t\sqrt{u/\nu}} f_Z(z)f_U(u)dzdu \end{aligned}$$

I can differentiate this with respect to t by simply differentiating the inner integral:

$$\frac{\partial}{\partial t}\int_{at}^{bt} f(x)dx = bf(bt) - af(at)$$

by the fundamental theorem of calculus. Hence

$$\frac{d}{dt}P(T \leq t) = \int_0^\infty f_U(u)\sqrt{u/\nu}\frac{\exp[-t^2u/(2\nu)]}{\sqrt{2\pi}}du\,.$$

Now I plug in

$$f_U(u) = \frac{1}{2\Gamma(\nu/2)}(u/2)^{(\nu-2)/2}e^{-u/2}$$

to get

$$f_T(t) = \int_0^\infty \frac{1}{2\sqrt{\pi\nu}\Gamma(\nu/2)}(u/2)^{(\nu-1)/2}\exp[-u(1+t^2/\nu)/2]\,du\,.$$

Make the substitution $y = u(1+t^2/\nu)/2$, $dy = (1+t^2/\nu)du/2$ $(u/2)^{(\nu-1)/2} = [y/(1+t^2/\nu)]^{(\nu-1)/2}$ to get

$$f_T(t) = \frac{1}{\sqrt{\pi\nu}\Gamma(\nu/2)}(1+t^2/\nu)^{-(\nu+1)/2}\int_0^\infty y^{(\nu-1)/2}e^{-y}dy$$

or

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