# STAT 801 Lecture 4

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

- Do 1 sample normal distribution theory.
- Define expectation, moments.
- Introduce moment generating function.

**Review of end of last time**

**Theorem 1** *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**: Reduces to $\mu = 0$ and $\sigma = 1$.
**Step 1**: Define

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

for suitable matrix $M$. Then

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

where $\mathbf{y}_2 = (y_2, \ldots, y_n)^t$.

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 (remember that $Z_n - \bar{Z} = -\sum_1^{n-1}(Z_i - \bar{Z})$).

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 $\sqrt{n}\bar{Z} \sim N(0,1)$.

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 $\chi^2$ density.

Suppose that $Z_1, \ldots, Z_n$ are independent $N(0,1)$. We define the $\chi^2_n$ 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 $\theta$ values run from 0 to $\pi$ except for the last $\theta$ whose values run from 0 to $2\pi$.) 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 $\theta_1$. 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 $u^{(n-1)/2}/2$ 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

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

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

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

so that the $\chi^2$ 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 $t_\nu$ 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:

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

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

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

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

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

## Expectation, moments

In elementary courses we give two definitions of expected values:

**Def'n** If $X$ has density $f$ then

$$E(g(X)) = \int g(x)f(x)\,dx\,.$$

**Def'n**: If $X$ has discrete density $f$ then

$$E(g(X)) = \sum_x g(x)f(x)\,.$$

Now if $Y = g(X)$ for smooth $g$ then

$$E(Y) = \int yf_Y(y) = \int g(x)f_Y(g(x))g'(x)\,dy = E(g(X))$$

by the change of variables formula for integration. This is good because otherwise we might have two different values for $E(e^X)$.

In general, there are random variables which are neither absolutely continuous nor discrete. Here's how probabilists define $E$ in general.

**Def'n**: A random variable $X$ is simple if we can write

$$X(\omega) = \sum_1^n a_i 1(\omega \in A_i)$$

for some constants $a_1, \ldots, a_n$ and events $A_i$.

**Def'n**: For a simple rv $X$ we define

$$E(X) = \sum a_i P(A_i)$$

For positive random variables which are not simple we extend our definition by approximation:

**Def'n**: If $X \ge 0$ then

$$E(X) = \sup\{E(Y) : 0 \le Y \le X, Y \text{ simple}\}$$

**Def'n**: We call $X$ **integrable** if

$$E(|X|) < \infty \,.$$

In this case we define

$$E(X) = E(\max(X, 0)) - E(\max(-X, 0))$$

Facts: $E$ is a linear, monotone, positive operator:

1. **Linear**: $E(aX + bY) = aE(X) + bE(Y)$ provided $X$ and $Y$ are integrable.

2. **Positive**: $P(X \ge 0) = 1$ implies $E(X) \ge 0$.

3. **Monotone**: $P(X \ge Y) = 1$ and $X$, $Y$ integrable implies $E(X) \ge E(Y)$.

Major technical theorems:

**Monotone Convergence**: If $0 \le X_1 \le X_2 \le \cdots$ and $X = \lim X_n$ (which has to exist) then

$$E(X) = \lim_{n\to\infty} E(X_n)$$

**Dominated Convergence**: If $|X_n| \le Y_n$ and there is a random variable $X$ such that $X_n \to X$ (technical details of this convergence later in the course) and a random variable $Y$ such that $Y_n \to Y$ with $E(Y_n) \to E(Y) < \infty$ then

$$E(X_n) \to E(X)$$

This is often used with all $Y_n$ the same random variable $Y$.

**Fatou's Lemma**: If $X_n \ge 0$ then

$$E(\limsup X_n) \le \limsup E(X_n)$$

**Theorem**: With this definition of $E$ if $X$ has density $f(x)$ (even in $R^p$ say) and $Y = g(X)$ then

$$E(Y) = \int g(x) f(x) dx \, .$$

(This could be a multiple integral.) If $X$ has pmf $f$ then

$$E(Y) = \sum_x g(x) f(x) \, .$$

This works for instance even if $X$ has a density but $Y$ doesn't.

**Def'n**: The $r^{\text{th}}$ moment (about the origin) of a real random variable $X$ is $\mu_r' = E(X^r)$ (provided it exists). We generally use $\mu$ for $E(X)$. The $r^{\text{th}}$ central moment is

$$\mu_r = E[(X-\mu)^r]$$

We call $\sigma^2 = \mu_2$ the variance.

**Def'n**: For an $R^p$ valued random vector $X$ we define $\mu_X = E(X)$ to be the vector whose $i^{\text{th}}$ entry is $E(X_i)$ (provided all entries exist).

**Def'n**: The $(p \times p)$ variance covariance matrix of $X$ is

$$Var(X) = E\left[(X-\mu)(X-\mu)^t\right]$$

which exists provided each component $X_i$ has a finite second moment.

Moments and probabilities of rare events are closely connected as will be seen in a number of important probability theorems. Here is one version of Markov's inequality (one case is Chebyshev's inequality):

$$\begin{aligned} P(|X-\mu| \ge t) &= E[1(|X-\mu| \ge t)] \\ &\le E\left[\frac{|X-\mu|^r}{t^r} 1(|X-\mu| \ge t)\right] \\ &\le \frac{E[|X-\mu|^r]}{t^r} \end{aligned}$$

The intuition is that if moments are small then large deviations from average are unlikely.

**Example moments**: If $Z$ is standard normal then

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} z e^{-z^2/2} dz/\sqrt{2\pi} \\ &= \left. \frac{-e^{-z^2/2}}{\sqrt{2\pi}} \right|_{-\infty}^{\infty} \\ &= 0 \end{aligned}$$

and (integrating by parts)

$$\begin{aligned} E(Z^r) &= \int_{-\infty}^{\infty} z^r e^{-z^2/2} dz/\sqrt{2\pi} \\ &= \left. \frac{-z^{r-1} e^{-z^2/2}}{\sqrt{2\pi}} \right|_{-\infty}^{\infty} + (r-1) \int_{-\infty}^{\infty} z^{r-2} e^{-z^2/2} dz/\sqrt{2\pi} \end{aligned}$$

so that

$$\mu_r = (r-1)\mu_{r-2}$$

for $r \geq 2$. Remembering that $\mu_1 = 0$ and

$$\mu_0 = \int_{-\infty}^{\infty} z^0 e^{-z^2/2} dz/\sqrt{2\pi} = 1$$

we find that

$$\mu_r = \begin{cases} 0 & r \text{ odd} \\ (r-1)(r-3)\cdots 1 & r \text{ even} \end{cases}$$

If now $X \sim N(\mu, \sigma^2)$, that is, $X \sim \sigma Z + \mu$, then $E(X) = \sigma E(Z) + \mu = \mu$ and

$$\mu_r(X) = E[(X-\mu)^r] = \sigma^r E(Z^r)$$

In particular, we see that our choice of notation $N(\mu, \sigma^2)$ for the distribution of $\sigma Z + \mu$ is justified; $\sigma$ is indeed the variance.

### Moments and independence

**Theorem**: If $X_1, \ldots, X_p$ are independent and each $X_i$ is integrable then $X = X_1 \cdots X_p$ is integrable and

$$E(X_1 \cdots X_p) = E(X_1) \cdots E(X_p)$$

**Proof**: Suppose each $X_i$ is simple: $X_i = \sum_j x_{ij} 1(X_i = x_{ij})$ where the $x_{ij}$ are the possible values of $X_i$. Then

$$\begin{aligned}
E(X_1 \cdots X_p) &= \sum_{j_1 \ldots j_p} x_{1j_1} \cdots x_{pj_p} E(1(X_1 = x_{1j_1}) \cdots 1(X_p = x_{pj_p}) \\
&= \sum_{j_1 \ldots j_p} x_{1j_1} \cdots x_{pj_p} P(X_1 = x_{1j_1} \cdots X_p = x_{pj_p}) \\
&= \sum_{j_1 \ldots j_p} x_{1j_1} \cdots x_{pj_p} P(X_1 = x_{1j_1}) \cdots P(X_p = x_{pj_p}) \\
&= \left[ \sum_{j_1} x_{1j_1} P(X_1 = x_{1j_1}) \right] \cdots \left[ \sum_{j_p} x_{pj_p} P(X_p = x_{pj_p}) \right] \\
&= \prod E(X_i)
\end{aligned}$$

For general $X_i > 0$ we create a sequence of simple approximations by rounding $X_i$ down to the nearest multiple of $2^{-n}$ (to a maximum of $n$) and applying the case just done and the monotone convergence theorem. The general case uses the fact that we can write each $X_i$ as the difference of its positive and negative parts:

$$X_i = \max(X_i, 0) - \max(-X_i, 0)$$

### Moment Generating Functions

**Def'n**: The moment generating function of a real valued $X$ is

$$M_X(t) = E(e^{tX})$$

defined for those real $t$ for which the expected value is finite.

**Def'n**: The moment generating function of $X \in R^p$ is

$$M_X(u) = E[\exp u^t X]$$

defined for those vectors $u$ for which the expected value is finite.

The mgf has the following formal connection to moments:

$$\begin{aligned} M_X(t) &= \sum_{k=0}^{\infty} E[(tX)^k]/k! \\ &= \sum_{k=0}^{\infty} \mu_k' t^k/k! \end{aligned}$$

It is thus sometimes possible to find the power series expansion of $M_X$ and read off the moments of $X$ from the coefficients of the powers $t^k/k!$.

**Theorem**: If $M$ is finite for all $t \in [-\epsilon, \epsilon]$ for some $\epsilon > 0$ then

1. Every moment of $X$ is finite.
2. $M$ is $C^\infty$ (in fact $M$ is analytic).
3. $\mu_k' = \frac{d^k}{dt^k} M_X(0)$.

The proof, and many other facts about mgfs, rely on techniques of complex variables.