18.443, Feb. 6, 2008

FACTS ABOUT NORMAL DISTRIBUTIONS AND SAMPLE STATISTICS

First, here is a known fact about normal distributions.

Theorem 1. If X and Y are independent random variables with normal distributions, $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$ then $X+Y$ is also normal, with $X+Y \sim N(\mu+\nu, \sigma^2+\tau^2)$.

This is proved in the "addnormals.pdf" handout posted on the course website. Paper copies aren't being distributed in class because we assume many of you know this fact from a probability course.

The next fact is stated early in Section 6.3 of Rice, p. 195. For any $X_1, ..., X_n$, $\overline{X}$ is defined as the sample mean $\overline{X} := (X_1 + \cdots + X_n)/n$.

Theorem 2. Let $X_1, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$. Then $\overline{X} \sim N(\mu, \sigma^2/n)$.

Proof. For any distribution F having finite mean μ and variance σ^2, if $X_1, \ldots, X_n$ are i.i.d. (F), then $\overline{X}$ has mean μ and variance σ^2/n. So the only problem is to show that $\overline{X}$ has a normal distribution in this case. Now, S_n defined as $X_1 + \cdots + X_n$ has a normal distribution, specifically $N(n\mu, n\sigma^2)$, by Theorem 1 and induction. Multiplying by a constant $1/n$ gives $\overline{X}$ which then has the stated distribution, Q.E.D.

In statistics, the mean μ and variance σ^2 of a distribution may be unknown and can be estimated from the data by the sample mean $\overline{X}$ and sample variance

$$s_X^2 \;=\; \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2,$$

defined for $n \geq 2$, respectively. The next fact includes Corollary A and Theorem B in Section 6.3 of Rice. It gives the distribution of s_X^2 (depending on σ^2) and its independence of $\overline{X}$ in the normal case. Rice uses the notation S^2 instead of s_X^2.

Theorem 3. If $X_1, ..., X_n$ are i.i.d. $N(\mu, \sigma^2)$, $n \geq 2$, then
(a) $\overline{X}$ and s_X^2 are independent random variables;
(b) $(n-1)s_X^2/\sigma^2$ has a $\chi^2(n-1)$ distribution.

Proof. Let $Y_j = X_j - \mu$ for $j = 1, \ldots, n$. Then $\overline{Y} = \overline{X} - \mu$ and $s_Y^2 = s_X^2$. So we can assume $\mu = 0$.

It's convenient to make a rotation of coordinates in n-space. Let the standard basis vectors be $\delta_i = \{\delta_{ij}\}_{j=1}^n$ where $\delta_{ij} = 1$ for $i = j$ and 0 for $i \neq j$. The first element of the new basis will be $e_1 = (1/\sqrt{n}, \ldots, 1/\sqrt{n})$. This does have length 1. Then we can always find further orthonormal basis vectors $e_2, ..., e_n$, for example $e_2 = (1/\sqrt{2}, -1/\sqrt{2}, 0, \ldots, 0)$, $e_3 = (1/\sqrt{6}, 1/\sqrt{6}, -2/\sqrt{6}, 0, \ldots, 0)$, etc.

For any two vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ (with respect to the standard basis) we have the usual dot product $x \cdot y = \sum_{j=1}^n x_j y_j$, with the squared length of x given by $|x|^2 = x \cdot x$.

Now, for the random vector $X = (X_1, \ldots, X_n)$ we have $\overline{X} = X \cdot e_1/\sqrt{n}$, and $(\overline{X}, \ldots, \overline{X}) = (X \cdot e_1)e_1$, which is the projection of X to the e_1 axis. The lengths of vectors and their dot products are preserved by rotations of coordinates, so

$$\sum_{j=1}^{n}(X_j - \overline{X})^2 \;=\; |X - (X \cdot e_1)e_1|^2 \;=\; \sum_{i=2}^{n}(X \cdot e_i)^2.$$

Recall that $\exp(y)$ is a notation for e^y. Since $X_1, \ldots, X_n$ are i.i.d. $N(0, \sigma^2)$, their joint density is

$$(\sigma\sqrt{2\pi})^{-n}\Pi_{j=1}^{n} \exp(-x_j^2/(2\sigma^2)) \;=\; (\sigma\sqrt{2\pi})^{-n} \exp(-|x|^2/(2\sigma^2)).$$

This distribution is invariant under any rotation of coordinates (change of orthonormal basis), specifically $|x|^2 = (x \cdot e_1)^2 + (x \cdot e_2)^2 + \cdots (x \cdot e_n)^2$. Thus $X \cdot e_1, \ \ldots, X \cdot e_n$ are i.i.d. $N(0, \sigma^2)$ and $X \cdot e_i/\sigma$ are i.i.d. $N(0, 1)$. It follows that $\overline{X} = X \cdot e_1/\sqrt{n}$ is independent of $s_X^2 = (n-1)^{-1}\sum_{i=2}^{n}(X \cdot e_i)^2$, proving (a). Also, $(n-1)s_X^2/\sigma^2 = \sum_{i=2}^{n}(X \cdot e_i)^2/\sigma^2$ has a $\chi^2(n-1)$ distribution, proving (b), Q.E.D.

Here is another way of looking at chi-squared distributions. As noted in the above proof, if $X_1, ..., X_d$ are i.i.d. $N(0, 1)$, their joint density is $(2\pi)^{-d/2}\exp(-|x|^2/2)$ on d-dimensional space. Let $Y = X_1^2 + \cdots + X_d^2$, so that Y has a $\chi^2(d)$ distribution. We have $P(Y \leq t) = 0$ for $t \leq 0$. For $t > 0$, $P(|Y| \leq t)$ is the integral of the density over the region where $|x|^2 \leq t$, or equivalently $|x| \leq \sqrt{t}$. Suppose $d \geq 2$. Using spherical coordinates, the integral becomes $A_d(2\pi)^{-d/2}\int_0^{\sqrt{t}} r^{d-1}\exp(-r^2/2)dr$ where A_d is a constant depending on d, the $(d-1)$-dimensional surface area of the unit sphere $|x| = 1$ in d-space. By the substitution $x = r^2$, $r = \sqrt{x}$, $dr = dx/(2\sqrt{x})$, the integral becomes

$$A_d(2\pi)^{-d/2}\int_0^t x^{(d-2)/2}\exp(-x/2)dx/2.$$

Since $(d-2)/d = (d/2) - 1$, and a probability density has a unique normalizing constant, this gives another proof that the $\chi^2(d)$ distribution is the $\Gamma(d/2, 1/2)$ distribution. Moreover, since we know that the normalizing constant is $(1/2)^{d/2}/\Gamma(d/2)$, we can evaluate $A_d = 2\pi^{d/2}/\Gamma(d/2)$. For example, if $d = 2$, since $\Gamma(1) = 0! = 1$, we get $A_2 = 2\pi$, the circumference of the unit circle as desired. If $d = 3$, then by the recursion formula, $\Gamma(3/2) = \Gamma(1/2)/2 = \sqrt{\pi}/2$, so $A_3 = 4\pi$, which is in fact the area of the unit sphere in 3 dimensions. Also, the volume of the unit ball $\{|x| \leq 1\}$ in d dimensions is $V_d = A_d\int_0^1 r^{d-1}dr = A_d/d = \pi^{d/2}/\Gamma((d/2) + 1)$, giving $V_2 = \pi$ and $V_3 = 4\pi/3$ as desired.