Handout 10: Math/Stat 394: Probability I

The Central Limit Theorem (CLT)

Wellner; 3/3/2000

Independent Repetitions: (Sampling With Replacement is a Special Case).
- The basic X experiment has mean μ and standard deviation σ.
- Let $T_n \equiv X_1 + \cdots + X_n$ for independent repetitions $X_1, \ldots, X_n$.
- T_n has mean $n\mu$ and standard deviation $\sqrt{n}\sigma$.
- $\overline{X}_n$ has mean μ and standard deviation $\sigma/\sqrt{n}$.
- The standardized random variable

$$Z_n \equiv \frac{T_n - n\mu}{\sqrt{n}\sigma} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

has mean 0 and standard deviation 1.

CLT: The distribution of Z_n converges to the $N(0,1)$ distribution: for any real numbers $a < b$

$$P(a < Z_n < b) \rightarrow P(a < Z < b)$$

where $Z \sim N(0,1)$.

Sampling Without Replacement: from an $a_1, \ldots, a_N$ urn with $\mu = \overline{a} = \sum_1^N a_i/N$ and $\sigma = \sigma_a^2 = N^{-1}\sum_1^N (a_i - \overline{a})^2$.
- One draw from the urn has mean $\mu = \overline{a}$ and standard deviation σ_a.
- Let $T_n \equiv X_1 + \cdots + X_n$ for the dependent repetitions $X_1, \ldots, X_n$.
- T_n has mean $n\overline{a}$ and standard deviation $\sqrt{n}\sigma_a\sqrt{1-(n-1)/(N-1)}$.
- $\overline{X}_n$ has mean $\overline{a}$ and standard deviation $(\sigma_a/\sqrt{n})\sqrt{1-(n-1)/(N-1)}$.
- The standardized random variable

$$Z_n \equiv \frac{T_n - n\overline{a}}{\sqrt{n}\sigma_a\sqrt{1-(n-1)/(N-1)}} = \frac{\overline{X}_n - \overline{a}}{(\sigma/\sqrt{n})\sqrt{1-(n-1)/(N-1)}}$$

has mean 0 and standard deviation 1.

Finite Sampling CLT: The distribution of Z_n converges to the $N(0,1)$ distribution provided $n \rightarrow \infty$ and $N - n \rightarrow \infty$ and σ_a^2 does not converge to 0 as $N \rightarrow \infty$: for any real numbers $a < b$

$$P(a < Z_n < b) \rightarrow P(a < Z < b)$$

where $Z \sim N(0,1)$.

Convolution formulas: Suppose that X, Y are independent.
Let $T \equiv X + Y$.
Then

$$\begin{array}{ll} p_T(t) = \sum_{\text{all } x} p_X(x) p_Y(t-x) & \text{or } = \sum_{\text{all } y} p_Y(y) p_X(t-y) \text{ (discrete case)} \\ f_T(t) = \int_{\text{all } x} f_X(x) f_Y(t-x) dx & \text{or } = \int_{\text{all } y} f_Y(y) f_X(t-y) dy \text{ (continuous case)} \end{array}$$

Proof: Discrete case:

$$\begin{aligned} p_T(t) &= P(T=t) = \sum_x P(X=x, T=X+Y=t) = \sum_x P(X=x, Y=t-x) \\ &= \sum_x P(X=x) P(Y=t-x) \qquad \text{by independence of } X, Y \\ &= \sum_x p_X(x) p_Y(t-x) \,. \end{aligned}$$

Continuous case: in this case we first compute the distribution of T Using the independence of X, Y it follows that the joint density of X, Y is given by the product of the marginal densities: $f_{X,Y}(x,y) = f_X(x) f_Y(y)$. Then

$$\begin{aligned} F_T(t) &= P(T \le t) = P(X+Y \le t) = \int\int_{(x,y):x+y\le t} f_X(x) f_Y(y) dx dy \\ &= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{t-x} f_X(x) f_Y(y) dy dx \\ &= \int_{x=-\infty}^{\infty} f_X(x) \left(\int_{y=-\infty}^{t-x} f_Y(y) dy \right) dx \\ &= \int_{x=-\infty}^{\infty} f_X(x) F_Y(t-x) dx \,. \end{aligned}$$

Now by differentiating across this identity and interchanging the derivative and the integral on the right side we find that

$$f_T(t) = F_T'(t) = \int_{-\infty}^{\infty} f_X(x) F_Y'(t-x) dx = \int_{-\infty}^{\infty} f_X(x) f_Y(t-x) dx \,.$$

□

Normal Distribution Facts:
A. If $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$ are independent, then
$X + Y \sim N(\mu + \nu, \sigma^2 + \tau^2)$.
B. If $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, then
- $T_n \equiv X_1 + \cdots + X_n \sim N(n\mu, n\sigma^2)$, and
- $\overline{X}_n \equiv n^{-1} T_n \sim N(\mu, \sigma^2/n)$.

The Chi-Square Distribution:
A. If $Z \sim N(0,1)$, then
$X \equiv Z^2 \sim$ Gamma$(1/2, 1/2)$ =Chi-square(1) with density

$$f_X(x) = \frac{(x/2)^{-1/2}}{2\sqrt{\pi}} e^{-x/2} 1_{(0,\infty)}(x).$$

B. If $Z_1, \cdots, Z_r$ are independent $N(0,1)$, then
$T_r \equiv Z_1^2 + \cdots + Z_r^2 \sim$ Gamma$(r/2, 1/2)$ =Chi-square(r)
with density

$$f_{T_r}(x) = \frac{(x/2)^{r/2-1}}{2\Gamma(r/2)} e^{-x/2} 1_{(0,\infty)}(x) \,.$$

C. If $X = Z^2$ with $Z \sim N(0,1)$ (so that $X \sim$ Chi-square(1)), then $E(X) = E(Z^2) = 1$ and $Var(X) = E(Z^4) - (E(Z^2))^2 = 3 - 1 = 2$. Thus for $T_r \sim$ Chi-square(r), $E(T_r) = r$ and $Var(T_r) = 2r$.

Proof of A and B:

$$\begin{aligned} F_X(x) &= P(X \le x) = P(Z^2 \le x) = P(-\sqrt{x} \le Z \le \sqrt{x}) \\ &= 2(P(Z \le \sqrt{x}) - P(Z \le 0)) = 2\Phi(\sqrt{x}) - 1 \,. \end{aligned}$$

Thus for $x > 0$,

$$f_X(x) = 2\phi(\sqrt{x})(1/2)x^{-1/2} = \frac{x^{-1/2}}{\sqrt{2\pi}} e^{-x/2} \,.$$

This is the Gamma$(1/2, 1/2)$ density, and this gets the special name Chi-square(1). B follows from A and the duplication property of the Gamma distribution. The only thing new in C is $E(Z^4) = 3$. But we have

$$\begin{aligned} E(Z^4) &= \int_{-\infty}^{\infty} z^4 \phi(z) dz = 2 \int_0^\infty \frac{z^4}{\sqrt{2\pi}} e^{-z^2/2} dz \\ &= 2 \int_0^\infty \frac{(2t)^{3/2}}{\sqrt{2\pi}} e^{-t} dt \end{aligned}$$

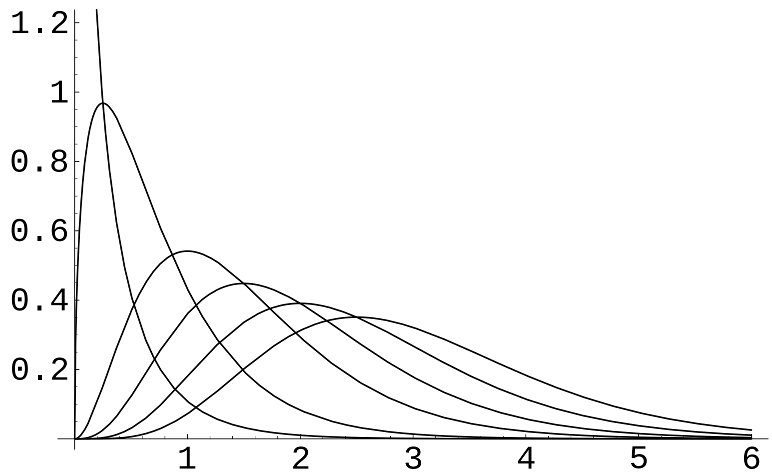

Figure 1: Plot of Chi-square$(k, 1)$ densities $k = 1, 3, 6, 8, 10, 12$.

$$
\begin{aligned}
&= \frac{2^2}{\sqrt{\pi}} \int_0^\infty t^{5/2-1} e^{-t} dt \\
&= \frac{2^2}{\sqrt{\pi}} \Gamma(5/2) \\
&= 3
\end{aligned}
$$

using

$$\Gamma(5/2) = (3/2)\Gamma(3/2) = (3/2)(1/2)\Gamma(1/2) = (3/2)(1/2)\sqrt{\pi}\,;$$

see Kelly pages 490-491. $\square$

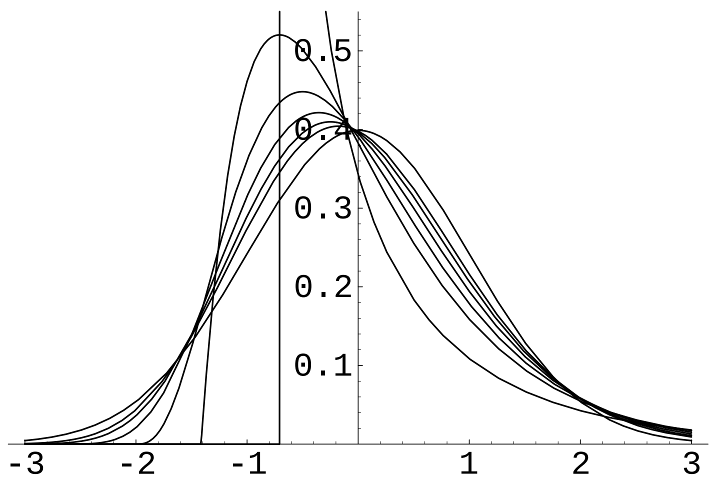

Figure 2: Plot of Standardized Chi-square$(k, 1)$ densities $k = 1, 4, 8, 16, 32, 64$.