Additional Notes for the Central Limit Theorem

This document illustrates the fact that the probability distribution for the sum of independent random quantities is the convolution of the distributions of those quantities. This is applied to determine the exact distribution of the total (and hence the sample mean) of a random sample of \( n \) exponentially distributed random quantities.

The exact p.d.f. can then be compared to a normal distribution with the same mean and variance, as the sample size \( n \) increases, thereby illustrating the central limit theorem.

Probability Distribution for the Sum of Independent Discrete Random Quantities

Let the random quantities \( X \) and \( Y \) be the scores on a pair of standard fair dice. Then they share the p.m.f. (probability mass function)

\[
p(x) = P[X = x] = \begin{cases} \frac{1}{6} & (x = 1, 2, 3, 4, 5, 6) \\ 0 & \text{otherwise} \end{cases}
\]

Let the random quantity \( T = X + Y \) be the total score on the two dice. We can construct the p.m.f. for \( T \).

As an example,

\[
P[T = 10] = P[X + Y = 10] = P[(X = 4 \cap Y = 6) \cup (X = 5 \cap Y = 5) \cup (X = 6 \cap Y = 4)]
\]

The three events in the union are all mutually exclusive. Therefore

\[
P[T = 10] = P[X = 4 \cap Y = 6] + P[X = 5 \cap Y = 5] + P[X = 6 \cap Y = 4]
\]

But events \( X = x \) and \( Y = y \) are independent. Therefore

\[
\]

or

\[
p_{x+y}(10) = p_x(4) \cdot p_y(6) + p_x(5) \cdot p_y(5) + p_x(6) \cdot p_y(4)
\]

In a similar way,

\[
p_{x+y}(11) = p_x(5) \cdot p_y(6) + p_x(6) \cdot p_y(5)
\]

\[
p_{x+y}(12) = p_x(6) \cdot p_y(6)
\]

\[
p_{x+y}(8) = p_x(2) \cdot p_y(6) + p_x(3) \cdot p_y(5) + p_x(4) \cdot p_y(4)
\]

\[
+ p_x(5) \cdot p_y(3) + p_x(6) \cdot p_y(2)
\]

\[
p_{x+y}(4) = p_x(1) \cdot p_y(3) + p_x(2) \cdot p_y(2) + p_x(3) \cdot p_y(1)
\]

and so on.
All of these summations for $P[X+Y = t]$ involve terms like $p_x(x) \cdot p_y(t-x)$, which is a **convolution sum**.

The lower limit of the summation is $x = (\text{the greater of } 1 \text{ and } (t-6))$.
The upper limit of the summation is $x = (\text{the lesser of } 6 \text{ and } (t-1))$.
The convolution sum for $P[X+Y = t]$ is therefore

$$P[X + Y = t] = \sum_{x = \max(1, t-6)}^{\min(6, t-1)} p_x(x) \cdot p_y(t-x)$$

More generally, for any two iid (independent and identically distributed) discrete random quantities $X_1, X_2$ with p.m.f. $P[X = x] = p(x)$, with $T = X_1 + X_2$,

$$P[T = t] = \sum_{x = \max(x_{\min}, t-x_{\max})}^{\min(x_{\max}, t-x_{\min})} p(x) \cdot p(t-x)$$

This result can be translated to the case of continuous random quantities.
Probability Distribution for the Sum of Ind’t Continuous Random Quantities

For any two iid continuous random quantities $X_1, X_2$ with p.d.f. $f(x)$, with $T = X_1 + X_2$,

$$f_T(t) = \int_{\min(\max(x_{\text{max}} - t, x_{\text{min}}))}^{\max(\min(x_{\text{max}} - t, x_{\text{min}}))} f(x) \cdot f(t-x) \, dx = f(x) \ast f(x)$$

If the p.d.f. of the random quantity has the same functional form $f(x)$ for all $x$, then $x_{\text{min}} = -\infty$, $x_{\text{max}} = +\infty$ and the convolution integral becomes

$$f_T(t) = \int_{-\infty}^{\infty} f(x) \cdot f(t-x) \, dx = f(x) \ast f(x)$$

It can be shown that the convolution of any normal distribution with itself is another normal distribution, with both mean and variance doubled.

If the p.d.f. of the random quantity has the functional form $f(x)$ for all positive $x$ only, (and is zero for $x < 0$), then $x_{\text{min}} = 0$, $x_{\text{max}} = +\infty$ and the convolution integral becomes

$$f_T(t) = \int_{0}^{t} f(x) \cdot f(t-x) \, dx = f(x) \ast f(x) \quad (t > 0)$$

(which is the familiar form for convolution integrals from Laplace transform theory).

On the next page it is shown that the convolution of any exponential distribution with itself is a gamma distribution $\Gamma(2, \lambda)$.

The sample mean $\overline{X}$ is the sample total divided by the sample size: $\overline{X} = \frac{T}{n}$.

Relationship between the probability density functions of $T$ and $\overline{X}$:

The c.d.f. for $T$ is $P[T \leq t] = F_T(t) = \int_{-\infty}^{t} f_T(u) \, du$

Apply a transformation of variables $u = nv \Rightarrow du = n \, dv$, $u = t \rightarrow v = t / n$:

$$P[T \leq t] = \int_{-\infty}^{t/n} n f_T(nv) \, dv \quad \text{But} \quad P[T \leq t] = P[\overline{X} \leq \frac{t}{n}] = \int_{-\infty}^{t/n} f_{\overline{X}}(v) \, dv$$

When two definite integrals over the same range of integration are equal, their integrands must be equal. Therefore

$$f_T(x) = n f_T(nx).$$
Sums of Exponentially Distributed Random Quantities

Let $X$ and $Y$ be iid (independent and identically distributed) random quantities, with the exponential distribution $f(x; \lambda) = \lambda e^{-\lambda x}, \ (x \geq 0, \lambda > 0)$.

Then the p.d.f. of the random quantity $T = X + Y$ is given by the convolution

$$f_T(t) = f(x)*f(x) = \int_0^t \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda(t-x)} \, dx \quad (t > 0)$$

$$= \lambda^2 \int_0^t e^{-\lambda(x+t-x)} \, dx = \lambda^2 e^{-\lambda t} \int_0^t 1 \, dx = \lambda^2 e^{-\lambda t} \left[ x \right]_0^t$$

Therefore the p.d.f. of $T = X + Y$ is the gamma distribution $\Gamma(2, \lambda)$:

$$f(t) = \lambda^2 t e^{-\lambda t} \quad (t > 0).$$

Gamma distributions for which the parameter $r$ is a natural number are also known as Erlang distributions (see Chapter 10). The general gamma p.d.f. is

$$f(x; r, \lambda) = \frac{x^{r-1} \lambda^r e^{-\lambda x}}{\Gamma(r)}, \quad (x \geq 0, \ r > 0, \ \lambda > 0)$$

For natural numbers $n$, $\Gamma(n) = (n-1)!$

We can also use convolution to find the p.d.f. of the sum of random quantities that follow Erlang distributions.

Let $X$ follow a gamma distribution with parameters $r = m$ and $\lambda$ and $Y$ follow a gamma distribution with parameters $r = n$ and $\lambda$, where $m$ and $n$ are natural numbers.

Then the p.d.f. of the random quantity $T = X + Y$ is given by the convolution

$$f_T(t) = f(x)*f(y) = \int_0^t \frac{\lambda^m x^{m-1} e^{-\lambda x}}{(m-1)!} \cdot \frac{\lambda^n (t-x)^{n-1} e^{-\lambda(t-x)}}{(n-1)!} \, dx \quad (t > 0)$$

$$= \frac{\lambda^{m+n} e^{-\lambda t}}{(m-1)! (n-1)!} \int_0^t x^{m-1} (t-x)^{n-1} \, dx$$

The integral is a standard Beta integral, which, through repeated integrations by parts, can be reduced to

$$f(t) = \frac{\lambda^{m+n} e^{-\lambda t}}{(m-1)! (n-1)!} \frac{(m-1)! (n-1)! t^{m+n-1}}{(m+n-1)!} = \frac{\lambda^{m+n} t^{m+n-1} e^{-\lambda t}}{(m+n-1)!}$$

which is the Erlang distribution $\Gamma(m+n, \lambda)$. 
Demonstration of the Central Limit Theorem (Exponential Distribution)

Therefore, if the random quantity $X$ follows an exponential distribution, with probability density function

$$f(x; \lambda) = \lambda e^{-\lambda x}, \quad (x \geq 0, \lambda > 0) \Rightarrow E[X] = \frac{1}{\lambda}, \quad V[X] = \frac{1}{\lambda^2}$$

then the sum $T$ of $n$ independent and identically distributed (iid) such random quantities follows the Erlang [Gamma] distribution, with p.d.f.

$$f(t; \lambda, n) = \frac{\lambda t^{n-1} e^{-\lambda t}}{(n-1)!}, \quad (t \geq 0, \lambda > 0, n \in \mathbb{N}) \Rightarrow E[T] = \frac{n}{\lambda}, \quad V[T] = \frac{n}{\lambda^2}$$

These $n$ values of $X$ form a random sample of size $n$.

Quoting $f_T(x) = n f_x(nx)$, the p.d.f. of the sample mean $\bar{X}$ is the related function

$$f_T(x; \lambda, n) = \frac{\lambda n (\lambda nx)^{n-1} e^{-\lambda nx}}{(n-1)!}, \quad (x \geq 0, \lambda > 0, n \in \mathbb{N})$$

and

$$E[\bar{X}] = \frac{1}{\lambda}, \quad V[\bar{X}] = \frac{1}{n \lambda^2}$$

For illustration, setting $\lambda = 1$, the p.d.f. for the sample mean for sample sizes $n = 1, 2, 4$ and 8 are:

- $n=1$: $f(x) = e^{-x}$
- $n=2$: $f_T(x) = 4xe^{-2x} = \frac{4(4x)^3 e^{-4x}}{3!}$
- $n=4$: $f_T(x) = \frac{4(4x)^3 e^{-4x}}{3!}$
- $n=8$: $f_T(x) = \frac{8(8x)^7 e^{-8x}}{7!}$

The population mean $\mu = E[X] = 1$ for all sample sizes.

The variance and the positive skew both diminish with increasing sample size.

The mode and the median approach the mean from the left.
For a sample size of \( n = 16 \), the sample mean \( \bar{X} \) has the p.d.f.\n\[
f_{\bar{X}}(x) = \frac{16(16x)^{15} e^{-16x}}{15!}
\] and parameters \( \mu = \text{E}[\bar{X}] = 1 \) and \( \sigma^2 = \text{V}[\bar{X}] = \frac{1}{16} \).

A plot of the exact p.d.f is drawn here, together with the normal distribution that has the same mean and variance. The approach to normality is clear. Beyond \( n = 40 \) or so, the difference between the exact p.d.f. and the Normal approximation is negligible.

It is generally the case (for distributions with well-defined mean and variance) that, whatever the probability distribution of a random quantity may be, the probability distribution of the sample mean \( \bar{X} \) approaches normality as the sample size \( n \) increases. For most probability distributions of practical interest, the normal approximation becomes very good beyond a sample size of \( n \approx 30 \).
Demonstration of the Central Limit Theorem (Binomial Distribution)

For probability distributions that are symmetric with one central peak to begin with, (such as binomial \((n, .5)\)), the approach to normality is even more rapid. As an illustration, the exact distribution for the mean of a random sample of size 4 taken from a binomial \((10, .5)\) distribution and the Normal distribution with the same mean and variance are plotted together here:

The agreement is remarkable, when one considers that the normal distribution is continuous but the binomial distribution is not.

For the binomial distribution, when the expected number of successes and the expected number of failures both exceed 10, the discrepancy between the exact and normal distributions is negligible even for a sample size of one!