Distributions of Functions of r.v.

Methods for finding the density function for a function of one or more random variables:

(1) The **CDF method**

- Let w be a function of random variables.
- Find the probability $P(w \leq w_0)$, which is (dropping the subscript 0) is equal to $F(w)$.
- $f_w(w) = dF(w)/dw$.

Example 7.3 Suppose the r.v. y has a density function

$$f_y(y) = \frac{e^{-y/\beta}}{\beta}, \quad 0 \leq y < \infty$$

and let $w(y) = y^2$. Find the density function for w.

(2) The **transformation method**

The density for r.v. y is known to be $f_y(y)$. Let w be a function (one-to-one) of y, i.e

$$w = h(y) \quad y = h^{-1}(w) = g(w)$$

Then, the density for w is equal to

$$f_w(w) = f_y(g(w))|\frac{dg(w)}{dw}|.$$

Why?

Example 7.3 Suppose the r.v. y has a density function

$$f_y(y) = \frac{e^{-y/\beta}}{\beta}, \quad 0 \leq y < \infty$$

and let $w(y) = y^2$. Find the density function for w.

Example 7.4 $x, y \sim U(0,1)$. Find the density function for the sum $w = x + y$.

- What's the range of w? – [0,2]
- Find the conditional density $f(w \mid x = x_0)$.

$$
\begin{aligned}
w = y + x_0 \quad &\rightarrow \quad y = w - x_0 \\
f_y(y) \quad &= \quad \mathbf{1}_{\{0 \le y \le 1\}} \\
f(w \mid x_0) \quad &= \quad \mathbf{1}_{\{0 \le w - x_0 \le 1\}}
\end{aligned}
$$

That is, $f(w \mid x) = \mathbf{1}_{\{x \le w \le x+1\}}$.

- Find the joint distribution for (w, x).

$$
\begin{aligned}
f(w, x) &= f(w \mid x) f(x) \\
&= \mathbf{1}_{\{x \le w \le x+1\}} \mathbf{1}_{\{0 \le x \le 1\}}
\end{aligned}
$$

- Find the marginal density for w.

$$f(w) = \int f(w, x) dx$$

When $w \in [0, 1]$,

$$f(w) = \int_0^w dx = w;$$

When $w \in (1, 2]$,

$$f(w) = \int_{w-1}^1 dx = 2 - w.$$

Distribution for CDF $F(y)$

Let y be a continuous r.v. with density function $f(y)$ and CDF $F(y)$.

Q: What's the distribution for $w = F(y)$?

A: $w \sim$ Uniform(0,1).

Using CDF method:

Simulation

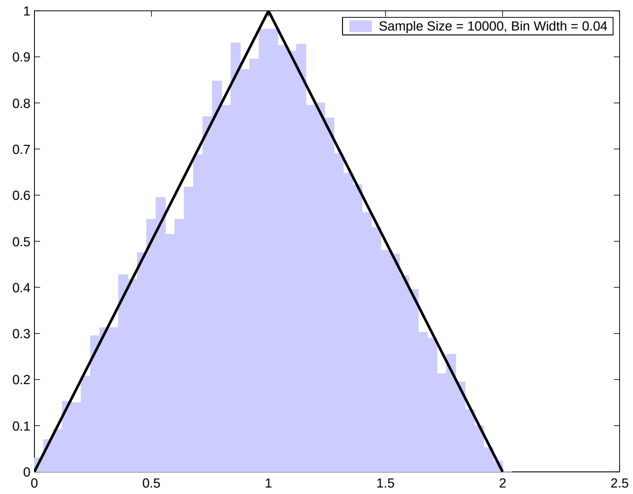
- How to generate *random data*?

- Approximate a sampling distribution by simulation

  (1) generate data from certain distribution

  (2) draw histogram with height equal to (relative freq / bin width)

  (3) the histogram can be used to approximate the density



- Monte Carlo Integration

$$\int h(t)f(t)dt = ?$$

Suppose $f(\cdot)$ is a density function, then

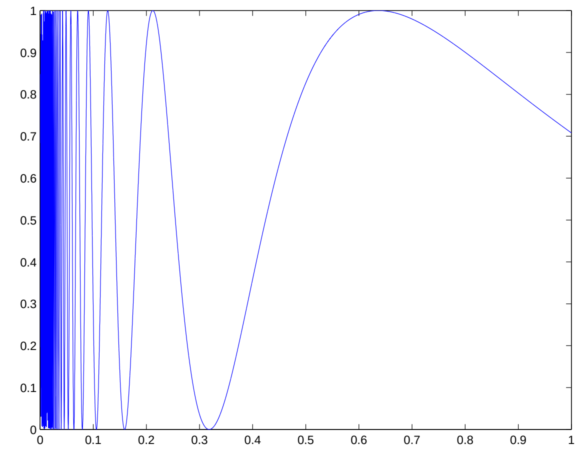
$$\int h(t)f(t)dt = \mathbb{E}h(y)$$

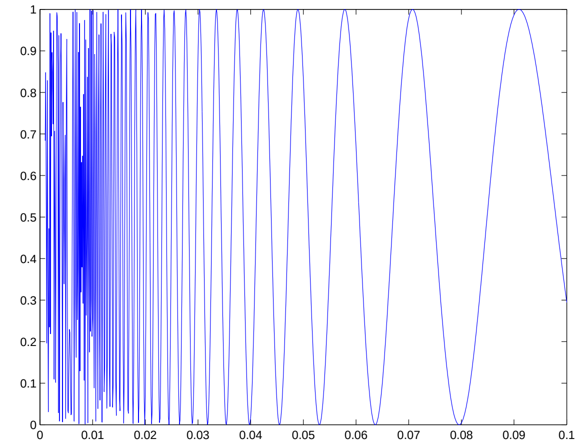
where $y \sim f(y)$. Then the expectation can be approximated by

$$\frac{1}{n}\sum_{i=1}^{n} h(y_i)$$

where $y_1, y_2 \ldots, y_n$ are random samples from the distribution $f(y)$.

Example : $\int_0^1 \sin(\frac{1}{x})^2 dx = ?$





- Mathematica – ans = 0.6735

- Matlab Code:

```
n=100000;

x = unifrnd(0, 1, n,1);

mean(sin(1./x).^ 2)
```

Outputs :

n=5000, ans = .6769

n=10000, ans = .6813

n=50000, ans= .6730

n=100000, ans = .6738

Sampling Distributions

Theorem 7.3 A linear combination of normally distributed random variables (even those that are correlated and have different means and variances), is normal distributed.

Example 7.8 Suppose we select independent random samples from two normal populations, n_1 from population 1 and n_2 from population 2.

If the means and variances for populations 1 and 2 are (μ_1, σ_1^2) and (μ_2, σ_2^2), respectively, and if $\bar{y}_1$ and $\bar{y}_2$ are the corresponding means.

What's the distribution of the difference $\bar{y}_1 - \bar{y}_2$.

- $y_1, \ldots y_n$ are drawn from a distribution **independently** with finite mean μ and variance σ^2, then

$$\mathbb{E}\bar{y} = \mu, \quad \mathrm{Var}(\bar{y}) = \sigma^2/n.$$

- **Central Limit Theorem**

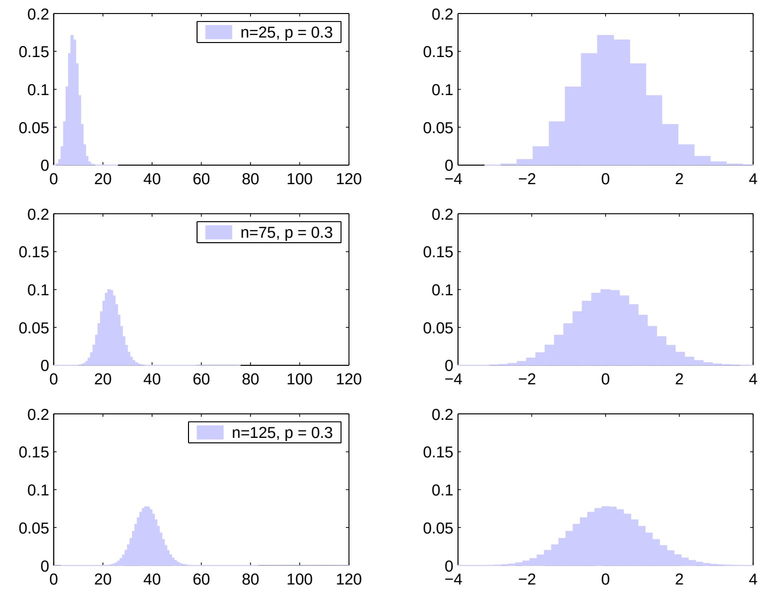
  When n is sufficiently large, $\bar{y}$ can be approximated by a normal distribution with mean μ and variance σ^2/n, i.e.,

$$\frac{\bar{y} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

- The sampling distribution of a sum of random variables

$$\sum_{i=1}^{n} y_i \sim N(n\mu, n\sigma^2)$$

Barplot for Binomial Distribution



Normal Approximation to Binomial

- $y \sim \text{Bi}(n, p)$, i.e.

$$y = \sum_{i=1}^{n} y_i, \quad y_i = 0, 1 \sim \text{Bernoulli}(p)$$

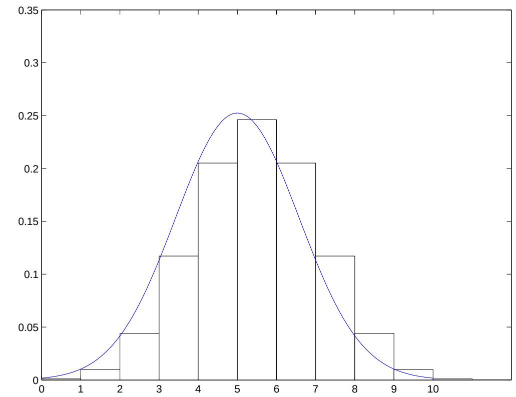
- $\mathbb{E}y = np, \quad \text{Var}(y) = npq$

- By Central Limiting Theorem,

$$\frac{y - np}{\sqrt{npq}} \sim N(0, 1)$$

$$P(y > np + x\sqrt{npq}) \approx \int_{\infty}^{x} \frac{\exp(-t^2/2)}{\sqrt{2\pi}} dt$$

- The approximation will be good if both $np \geq 4$ and $nq \geq 4$.



Continuity Correction for the Normal Approximation to a Binomial Probability

Let y be a $\text{Bin}(n, p)$ and let $z = (y - np)/\sqrt{npq}$. Then

$$P(y \leq a) \approx P(z < \frac{a + 0.5 - np}{\sqrt{npq}})$$

$$P(y \geq a) \approx P(z > \frac{a - 0.5 - np}{\sqrt{npq}})$$

$$P(a \leq y \leq b) \approx P(\frac{a - 0.5 - np}{\sqrt{npq}} < z < \frac{b + 0.5 - np}{\sqrt{npq}})$$

Sampling Dist Related to Normal

A random sample $y_1, y_2, \ldots, y_n$ is drawn from $N(\mu, \sigma^2)$.

$$\text{sample mean } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\text{sample var } s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

- $\chi^2 = (n-1)s^2/\sigma^2$ has a chi-square distribution with $(n-1)$ degrees of freedom.

  Why?

- If χ_1^2 and χ_2^2 are **independent** chi-square distribution with v_1 and v_2 degrees of freedom, then $\chi_1^2 + \chi_2^2$ has a chi-square distribution with $(v_1 + v_2)$ degrees of freedom.

  Why?

- Let x be a standard normal and χ^2 be a chi-square with v degrees of freedom. If x and z are independent, then

  $$t = \frac{z}{\sqrt{\chi^2/v}}$$

  has a **Student's t distribution** with v degree of freedom.

- Let χ_1^2 and χ_2^2 be independent chi-square distribution with v_1 and v_2 degrees of freedom respectively, then

  $$F = \frac{\chi_1^2/v_1}{\chi_2^2/v_2}$$

  has an **F distribution** with v_1 numerator degrees of freedom and v_2 denominator degrees of freedom.