Chapter 1: Probability theory and sampling

1.1. Basic probability theory

- Sample space : Ω or S — set/space of all possible outcomes
- Event : $E \subseteq \Omega$

Example : Flip coin once : $S = \{H, T\}$

" " twice : $S = \{HH, HT, TH, TT\}$

$\uparrow$ elementary event

- Probability function :
  - $\{p_i\}_{i=1}^{|S|}$
  - $p_i \geq 0 \quad i \in \Omega$
  - $\sum_{i=1}^{|S|} p_i = 1$
  - $P(E) = \sum_{i \in E} p_i$ $\qquad P(\Omega) = 1$

- Empty event : $\emptyset$, $P(\emptyset) = 0$
- Operations on/combination of events :

$P(A \cup B) = \text{Prob}(A \text{ or } B)$ $\qquad P(A \cup B) = P(B \cup A)$

$P(A \cap B) = \text{Prob}(A \text{ and } B) = P(A, B)$ $\qquad P(A \cap B) = P(B \cap A$

$= P(AB)$

$P(A^c) = 1 - P(A)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- De Morgan : $(E \cup F)^c = E^c \cap F^c$

$(E \cap F)^c = E^c \cup F^c$

- Mutually exclusive : E, F such that $E \cap F = \emptyset$

$\Rightarrow P(E \cup F) = P(E) + P(F)$

  - Note : $\emptyset \cap \emptyset = \emptyset$ $\qquad$ $\emptyset$ m.e. with itself

- Ref: GS Chap 1

L1, h1

GS: Sec 1.4 1.2. Conditional probabilities

$$P(E|F) = \frac{P(E\cap F)}{P(F)} = \frac{P(E,F)}{P(F)} \qquad P(F)>0$$

- Can't condition on event $F \ni P(F)=0$
- Interpretation #1: Prob of E given F happens
  or F observed
- Interpretation #2: Prob of E in subset of events in which F is satisfied
  ↳ constraint or restriction

· Multiplication rule:

$$P(E\cap F) = P(E|F)\,P(F)$$
$$= P(F|E)\,P(E)$$

$$P(E_1 \cap E_2 \cap \cdots) = P(E_1)\,P(E_2|E_1)\,P(E_3|E_1 E_2)\cdots$$

· Total probability:

$$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c)$$
$$\underbrace{P(E)}_{\text{marginal}} = \sum_i P(E|F_i)\,P(F_i) \qquad F_i \text{ mutually exclusive}$$

Decomposition of marginal over alternatives

· Bayes's formula/rule:

prior

$$\underbrace{P(F|E)}_{\text{posterior}} = \frac{P(E|F)\,P(F)}{P(E)}$$

F: event/hypothes
E: evidence

- Interpretation: Hypothesis → evidence → update
  $P(F)$ $P(E)$ $P(F|E)$
- General: $$P(F_j|E) = \frac{P(E|F_j)\,P(F_j)}{\sum_i P(E|F_i)\,P(F_i)}$$

· Independence: A, B independent if ($A \perp\!\!\!\perp B$)

- $P(A\cap B) = P(A,B) = P(A)\,P(B)$
- $P(A|B) = P(A)$
- $P(B|A) = P(B)$

not a Venn diagram property
not mutually exclusive

L1, h1

GS, Chap. 3

1.3. Discrete random variables (RVs)

- Def.: RV X defined by
  - Set of possible values
  - Probability for each value

- Notations: $X = x$ $\quad P(X=x)$ or $P\{X=x\}$ or $P(x)$ $\quad \sum_x P(x) = 1$

  RV value

- Example: Flip coin 3 times
  - Sample space: $\Omega = \{HHH, HHT, HTH, THH, \dots\}$
  - X = no. heads
  - $X \in \{0, 1, 2, 3\}$ $\quad P(0) = P(3) = \frac{1}{8}$ $\quad P(1) = P(2) = \frac{3}{8}$

- Expectation: $E[X] = \sum_x x\, P(X=x) = \sum_x x\, P(x)$
  - $E[a] = a$ $\qquad$ a constant
  - $E[X+Y] = E[X] + E[Y]$, $\qquad E[XY] = E[X]E[Y]$ if $X \perp\!\!\!\perp Y$
  - $E[aX+c] = a\,E[X] + c$ $\qquad$ a, c constants

- Variance: $\mathrm{var}(X) = E[(X - E[X])^2]$
  $= E[X^2] - E[X]^2 \geq 0$
  - $\mathrm{var}(aX+b) = a^2\, \mathrm{var}(X)$
  - $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$ if $X \perp\!\!\!\perp Y$

- Standard deviation: $\sigma(X) = \sqrt{\mathrm{var}(X)}$

- Bernoulli RV: $X \in \{0, 1\}$ $\quad p(0) = p$ $\quad P(1) = 1-p$

- Binomial RV:
  - Trial: success/failure $\quad$ 0/1 $\quad$ true/false $\quad$ H/T
  - X = # successes in n independent trials
  - $X \in \{0, 1, \dots, n\}$
  - $P(x) = \binom{n}{x} p^x (1-p)^{n-x}$ $\qquad X \sim \mathrm{Bin}(n,p)$
  - $E[X] = np$ $\qquad \mathrm{var}(X) = np(1-p)$

L1, h1,2

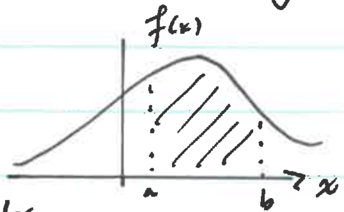
- Poisson RV:
  - $X \in \{0, 1, 2, \dots\}$
  - $P(X=i) = e^{-\lambda} \frac{\lambda^i}{i!}$ $\lambda > 0$
  - $E[X] = \lambda$ $\text{var}(X) = \lambda$

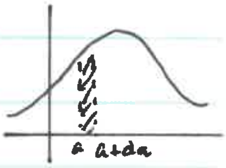
$X \sim \text{Poisson}(\lambda)$
limit of binomial
see CW1

GS, Chap. 4

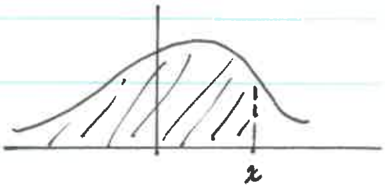
1.4 Continuous random variables

- Probability density function: $p_X(x)$ or $f_X(x)$ or $p(x)$ or $f(x)$
  - $f(x) \geq 0$
  - $\int_{-\infty}^{\infty} f(x)\,dx = 1$
  - $P(X \in [a,b]) = P(a \leq X \leq b) = \int_a^b f(x)\,dx$
  - $P(X \in A) = \int_A f(x)\,dx$
  - Interpretation: $P(X \in [a, a+da]) = f(a)\,da$





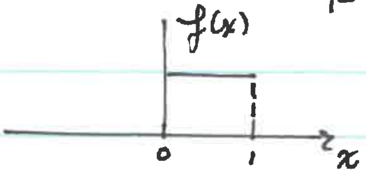
- Cumulative distribution function (CDF)
  $F(x) = P(X \leq x) = \int_{-\infty}^{x} p(y)\,dy$



- Expectation: $E[X] = \int_{-\infty}^{\infty} x\,p(x)\,dx$ same properties
- Variance: $\text{var}(X) = E[X^2] - E[X]^2$ "   "
- n^{th} moment: $E[X^n]$
- Joint pdf: $p_{XY}(x,y)$ $p_X(x) = \int p_{XY}(x,y)\,dy$
  $p_Y(y) = \int p_{XY}(x,y)\,dx$
  - $p_{XY}(x,y) = p_X(x)\,p_Y(x)$ if $X \perp\!\!\!\perp Y$

L1, L2

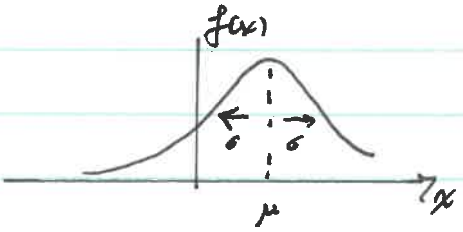
- Uniform RV : $X \sim \mathcal{U}[0,1]$
  - $X \in [0,1]$
  - $f(x) = \begin{cases} 1 & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$



- Exponential RV: $X \sim Exp(\lambda)$
  - $X \geq 0$, $X \in \mathbb{R}_+$
  - $f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

Normal

- Gaussian RV: $X \sim N(\mu, \sigma^2)$
  - $X \in \mathbb{R}$
  - $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ ~ normpdf(x, μ, σ^2) in Matlab
  - $E[X] = \mu$ $\quad$ $var(X) = \sigma^2$
  - Standardization : $\underset{Z}{Y} = \frac{(X-\mu)}{\sigma} \sim N(0,1)$ Standard normal
  - CDF : $\Phi(a) = P(Z \leq a)$



- Cauchy (Lorentzian) : $X \sim Cauchy(a, \sigma)$
  - $X \in \mathbb{R}$
  - $f(x) = \frac{1}{\pi} \frac{\sigma}{(x-a)^2 + \sigma^2}$
  - $E[X]$ undefined $\quad$ $var(X) = \infty$ (!?)

L1, L2

GS, Sec. 4.7 1.5. Transformation of RVs

Prop.:
- X continuous RV
- $Y = g(X)$
- g differentiable and monotonic

Then:

$$p_Y(y) = \begin{cases} p_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & y \ni g^{-1}(y) \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

- Mnemonic: $p_Y(y)\,dy = p_X(x)\,dx \Rightarrow p_Y(y) = p_X(x(y)) \frac{dx}{dy}$

- General:

$$p_Y(y) = \sum_{\underbrace{x \in g^{-1}(y)}_{\text{pre-image}}} p_X(g^{-1}(y)) \left| \frac{d\, g^{-1}(y)}{dy} \right|$$

$$P(y) = \sum_{x \in g^{-1}(y)} P(g^{-1}(y)) \qquad \text{for discrete RVs}$$

- Many dimensions / joint pdf:

$$p_{\vec{Y}}(\vec{y}) = p_{\vec{X}}(\vec{x}(\vec{y})) \underbrace{\left| \frac{\partial \vec{x}}{\partial \vec{y}} \right|}_{\text{Jacobian}}$$

- Example: $X \sim \mathcal{N}(\mu, \sigma^2)$ $\quad Y = aX + b$ $\quad$ linear transformation
  - $Y = g(X) = aX + b$
  - $X = g^{-1}(Y)$ $\qquad g^{-1}(y) = \frac{y-b}{a}$ $\qquad \frac{d}{dy} g^{-1}(y) = \frac{1}{a}$

$$\Rightarrow p_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{\left(\frac{y-b}{a} - \mu \right)^2}{2\sigma^2} \right) \frac{1}{|a|}$$

$$= \frac{1}{\sqrt{2\pi a^2\sigma^2}} \exp\left(- \frac{(y - a\mu - b)^2}{2\sigma^2 a^2} \right) = \frac{1}{\sqrt{2\pi\sigma'^2}} e^{-(x-\mu')^2 / 2\sigma'^2}$$

$\mu' = a\mu + b$

$\sigma'^2 = a^2\sigma^2$

Standardization: $Y = \frac{X-\mu}{\sigma}$

$\Rightarrow \mu' = 0$, $\sigma' = 1$

GS, Chap 5

1.6 Characteristic and generating functions

- Characteristic fct (CF):

$$G_X(k) = E[e^{ikX}] \qquad k \in \mathbb{R}$$

$$= \int_{-\infty}^{\infty} p_X(x)\, e^{ikx}\, dx \qquad \text{Fourier transform}$$

- (Moment) generating fct (GF):

$$M_X(k) = E[e^{kX}]$$

$$= \int_{-\infty}^{\infty} p_X(x)\, e^{kx}\, dx \qquad k \in \mathbb{R}$$

Laplace transform

- Cumulant function: $C_X(k) = \ln \begin{matrix} G_X(k) \\ M_X(k) \end{matrix}$

- Properties
  - $G_X(0) = M_X(0) = 1$
  - $X \perp\!\!\!\perp Y \Rightarrow E[e^{ik(X+Y)}] = E[e^{ikX} e^{ikY}] = E[e^{ikX}]\,E[e^{ikY}]$

    $\Rightarrow G_{X+Y}(k) = G_X(k)\, G_Y(k)$ or $M_{X+Y}(k)$
  - $M_X(k) = 1 + \sum_{n=1}^{\infty} \frac{k^n}{n!} \underbrace{E[X^n]}_{\text{moments}}$

- Example: $X \sim N(\mu, \sigma^2)$

$$\begin{tikzcd} G_X(k) = e^{ik\mu - \frac{\sigma^2}{2}k^2} \arrow[r, dashed, "\text{not always}"] & M_X(k) = e^{k\mu + \frac{\sigma^2}{2}k^2} \arrow[l, bend left, "k \to ik"] \end{tikzcd}$$

L2, h1

GS, Sec 5.10

1.7 Limit theorems

- Sequence of iid RVs:

  $X_1, X_2, \ldots, X_n$ $\quad$ X_i independent

  $X_i \sim p$ identically distributed

  $p(X_1, X_2, \ldots, X_n) = p(X_1)p(X_2)\cdots p(X_n)$

- Sum of RVs: $S_n = \sum_{i=1}^{n} X_i$

- Law of large numbers (weak):
  - $X_1, X_2, \ldots, X_n$ iid $X_i \sim p$
  - $\mu = E[X_i] < \infty$

Then: $\frac{S_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{n\to\infty} \mu$ in probability

i.e. $\lim_{n\to\infty} P\left(\left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) = 0$ $\qquad$ $\frac{S_n}{n}$ = sample mean

or

$\lim_{n\to\infty} P\left(\left| \frac{S_n}{n} - \mu \right| < \varepsilon \right) = 1$

- Central limit theorem:
  - $X_1, X_2, \ldots, X_n$ iid $X_i \sim p$
  - $\mathrm{var}(X_i) < \infty$

Then: $\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \right) = \mathcal{N}(0,1)$

Example: $X_1, X_2, \ldots, X_n$ iid $X_i \sim \mathcal{N}(\mu, \sigma^2)$

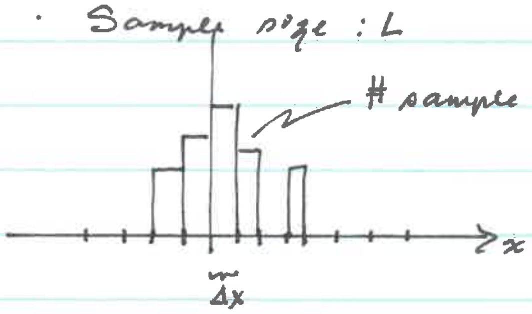
- $S_n \sim \mathcal{N}(n\mu, n\sigma^2)$ $\quad$ why?
- $\frac{S_n}{n} \sim \mathcal{N}(\mu, n\frac{\sigma^2}{n^2}) = \mathcal{N}(\mu, \frac{\sigma^2}{n}) \xrightarrow{n\to\infty} \delta(s-\mu)$
- $\frac{S_n - n\mu}{\sqrt{n}\,\sigma} \sim \mathcal{N}(0,1)$ standardization $\qquad$ CLT shows general property

- Ref: GS, Sec. 5.10 for proofs

12,11

1.8 Histograms

- $\{x_i\}_{i=1}^{L}$ sample (list, set) of L values (data)
- Sample size : L



sample values in $[x, x+\Delta x]$

- $\hat{N}_{L,\Delta x}(x) = \#$ samples in $[x, x+\Delta x]$ $\qquad \sum_x \hat{N}_{L,\Delta x}(x) = L$
- $\hat{P}_{L,\Delta x}(x) = \dfrac{\hat{N}_{L,\Delta x}(x)}{L}$ $\qquad \sum_x \hat{P}_{L,\Delta x}(x) = 1$
- $\hat{f}_{L,\Delta x}(x) = \dfrac{\hat{N}_{L,\Delta x}(x)}{\Delta x \; L} = \dfrac{\hat{P}_{L,\Delta x}(x)}{\Delta x}$ $\qquad \sum_x \hat{f}_{L,\Delta x}(x)\Delta x = 1$

$\approx \int f(x)\,dx = 1$

$$\begin{tikzcd}
\{X_i\} \arrow[r] & \hat{N}_{L,\Delta x} \arrow[r] & \hat{P}_{L,\Delta x} \arrow[r] & \hat{f}_{L,\Delta x} \\
\text{Data} & \text{Histogram count} & \text{Empirical distribution} & \text{Empirical density}
\end{tikzcd}$$

- LLN: If $x_i \sim p$ iid, then $\hat{f}_{L,\Delta x}(x) \xrightarrow{L\to\infty} p(x)$

- Code :

center points

plt.hist for plot

Matlab	hist	histc	
Python	np.histogram	np.bincount	
Mathematica	Histogram	HistogramList	BinCount
R	hist	binCounts	

- Own code : (Basic)

```
myhist(data, a, b, dx)
    L = size(data)
    nbin = (b-a)/dx
    hN = zeros(1, nbin)
    for i=1:L
        pos = $\lfloor \frac{data[i]-a}{dx} \rfloor$   or $\lceil \cdot \rceil$
        hN[pos]++
    end
    hN /= L dx
```

Incomplete: not test if data$\notin [a,b]$

LL, hL

1.9 Pseudo-random numbers

- $X \sim \mathcal{U}[0,1]$ Uniform random float

Matlab	rand	rand()	mxn matrix rand(m,n)
Mathematica	RandomReal[]		
Python	random.random()		Numpy random.random(size)
R	runif()		runif(size,a,b) $\mathcal{U}[a,b]$

- Example: Uniformity test

```
L = 10^4
dx = 0.1
vals = rand(1,L)
xvals = [0:dx:1]
counts = hist(vals,xvals)
hp = counts/(L*dx)
plot (xvals, hp)
plot (xvals, 1)
```

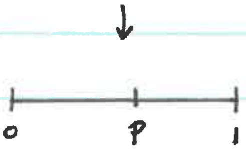
- Seed initialization:

Matlab	rng(seed)
Mathematica	SeedRandom[n]
Python	random.seed(n)
R	set.seed(n)

Seed always changes if not initialized

1.10 Non-uniform variates

· Method 1: Transformation of RVs



Example: $X \sim \mathcal{U}[0,1]$

$$Y = \mathbb{1}_{[0,p]}(X) = \begin{cases} 1 & \text{if } X \in [0,p] \\ 0 & \text{if } X \in [p,1] \end{cases}$$

$$\Rightarrow P(Y=1) = P(X \in [0,p]) = \int_0^p 1 \, dx = p \qquad \Rightarrow Y \sim \text{Bern}(p)$$

$$\Rightarrow P(Y=0) = 1-p$$

· Code:

```
y = Bern(p)
    r = rand()
    if  r < p
            y=1
    else
            y=0
    end
end
```

Coin flip

· Example: $X \sim \mathcal{N}(0,1)$. Generate $Y \sim \mathcal{N}(\mu, \sigma^2)$

$\Rightarrow$ Use $Y = \sigma X + \mu$

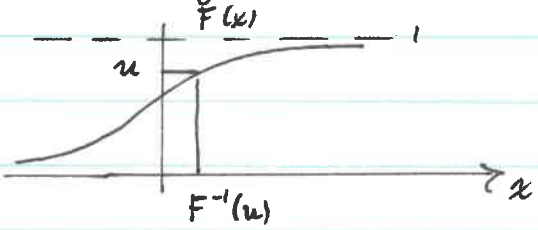
· Code:

```
x = randn()
y = σ*x + μ
```

Matlab	randn()	
Python	np.random.randn()	np.random.normal(μ,σ)

Demonstration

- Method 2: Inversion of CDF
  - X continuous RV
  - Cumulative distribution function (CDF): $F(x) = P(X \le x)$
    $= \int_{-\infty}^{x} f(y)\,dy$
  - Probability density: $f(x) = F'(x)$
  - Inverse of CDF: $F^{-1}(u) = x$ such that $F(x) = u$



$u \le F(x) \Leftrightarrow F^{-1}(u) \le x$

since $F(x)$ is monotonically increasing

- Proposition: If $U \sim \mathcal{U}[0,1]$, then $F^{-1}(U)$ has CDF F.

Proof:

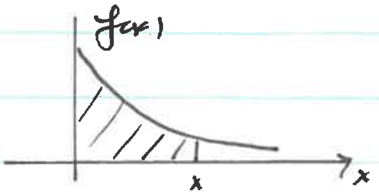
$$P(F^{-1}(U) \le x) = P(U \le F(x))$$
$$= F(x)$$

$P(U \le a) = a$ for uniform

▨

- Algorithm: ① Get CDF from PDF
  ② Invert CDF (not always possible analytically)
  ③ $u = \text{rand}()$.
  ④ $x = F^{-1}(u)$

- Example: Exponential distribution
  $f(x) = \lambda e^{-\lambda x}$ $x \ge 0$
  - CDF: $F(x) = P(X \le x)$
    $= \int_{-\infty}^{x} f(y)\,dy = 1 - e^{-\lambda x}$



- Inverse CDF: $u = F(x) = 1 - e^{-\lambda x} \Rightarrow x = F^{-1}(u) = -\frac{1}{\lambda} \ln(1-u)$
- $U \sim \mathcal{U}[0,1] \Rightarrow U \sim 1 - U$

$\Rightarrow$ can use $F^{-1}(u) = -\frac{1}{\lambda} \ln u$

- Example: Cauchy distribution
  - $f(x) = \frac{1}{\pi} \frac{\sigma}{x^2+\sigma^2}$
  - $F(x) = P(X \le x)$

    $= \int_{-\infty}^{x} f(y)\,dy = \int_{-\infty}^{x} \frac{1}{\pi}\frac{\sigma}{y^2+\sigma^2}\,dy$

    $= \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x}{\sigma}\right)$
  - $F^{-1}(u) = \sigma \tan\left(\underbrace{\pi\left(u-\frac{1}{2}\right)}_{\text{periodic with period 1}}\right)$

  $\Rightarrow$ Can use $F^{-1}(u) = \sigma\tan(\pi u)$

  $u = \text{rand}()$
  $X = \sigma\tan(\pi u)$

- Example: Pareto (if time)
  - $f(x) = \frac{ab^a}{x^{a+1}}$ $\quad 0 \le b \le x$
  - $F(x) = 1 - \left(\frac{b}{x}\right)^a$

    Used in finance
  - $F^{-1}(u) = \frac{b}{(1-u)^{1/a}}$ $\Rightarrow$ Can use $X = \frac{b}{U^{1/a}}$

- Example: Gaussian distribution
  - $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$
  - $F(x) = P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right)$

    $= P\left(Z \le \frac{x-\mu}{\sigma}\right)$

    $= \Phi\left(\frac{x-\mu}{\sigma}\right)$
  - $F^{-1}(u)$ involves Φ^{-1}. Not known in closed form

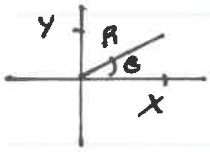
    Can be computed numerically

    Not efficient.

1.11 Box - Muller method

· Exercise in CW1 : $X \sim N(0,1)$ $\quad R = \sqrt{X^2+Y^2}$

$Y \sim N(0,1)$ $\quad \theta = \arctan \frac{Y}{X}$



$\Rightarrow$ R has Rayleigh distribution $\quad p(r) = r e^{-r^2/2}$

$\theta \sim \mathcal{U}[0, 2\pi)$

· Box-Muller :

1- $R \sim$ Rayleigh

2- $\theta \sim \mathcal{U}(0, 2\pi)$

output 3- $(X, Y) = (R \cos\theta, R \sin\theta)$ $\quad$ 2 standard normal RVs

· Step 1: $\quad F(r) = P(R \le r) = \int_0^r p(y)\,dy = 1 - e^{-r^2/2}$

$\Rightarrow F^{-1}(u) = \sqrt{-2\ln(1-u)}$

$\Rightarrow$ Choose $U_1 \sim \mathcal{U}[0,1]$

$R = \sqrt{-2\ln(1-U_1)}$ or $\sqrt{-2\ln U_1}$

· Step 2: $\quad U_2 \sim \mathcal{U}[0,1] \Rightarrow \theta = 2\pi U_2 \sim \mathcal{U}[0, 2\pi]$

· Remarks :

- Must choose/generate 2 uniform RVs
  Can't use U_1 for U_2
- Generate 2 Gaussians — output only 1.

1.12 Monte Carlo sampling

- RV : X
- Expectation : $\mu = E[X] = \sum_x x\, P(x)$ or $\int x\, p(x)\, dx$

  discrete   continuous
- General expectation: $\gamma = E[g(X)] = \sum_x g(x) P(x)$ or $\int g(x) p(x) dx$

- Monte Carlo method/estimation:
  - Generate samples $\{X_i\}_{i=1}^{L}$ $X_i \sim p(x)$ iid
  - Estimator:

  $$\hat{\gamma}_L = \frac{1}{L} \sum_{i=1}^{L} g(X_i)$$

  - LLN: $P(|\hat{\gamma}_L - \gamma| > \epsilon) \xrightarrow{L\to\infty} 0$

  $\hat{\gamma}_L \to \gamma$ in probability as $L \to \infty$
  - For $L \gg 1$, $\hat{\gamma}_L \approx \gamma$
  - Take $\hat{\gamma}_L$ as estimate of γ

- Unbiased estimator
- Maximum likelihood estimator

- Example: Expectation of Rayleigh distribution

  $p(r) = r\, e^{-r^2/2}$
  - Generate $\{r_i\}_{i=1}^{L}$ $r_i \sim$ Rayleigh
  - Estimator: $\hat{\mu}_L = \frac{1}{L}\sum_{i=1}^{L} r_i \longrightarrow E[R] = \int_0^\infty r\, p(r)\, dr$

  $$= \int_0^\infty r^2 e^{-r^2/2} dr = \sqrt{\frac{\pi}{2}}$$

```
L = 10^3
est = zeros(1,L)
s=0
for i=1:L
    x= randn(1,2)
    r= sqrt( x[1]^2 + x[2]^2 )
    s = s+r
  est(i)= s/i
end
plot( 1:L, est)
```

$2\hat{\mu}_L^2 \to \pi$

possible estimate of π

1.13 Statistical errors

- Estimator : $\hat{\mu}_L = \frac{1}{L}\sum_{i=1}^{L} X_i$ it's a RV

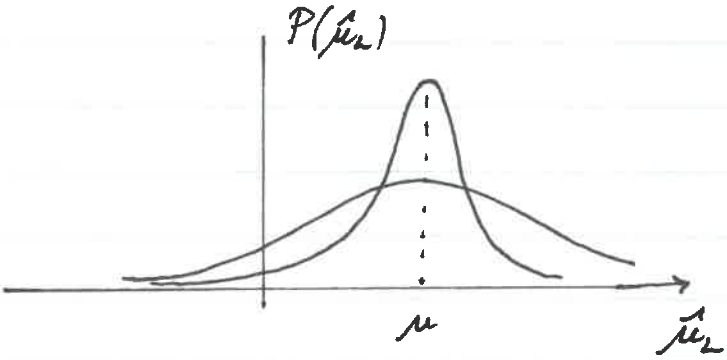
- Expectation : $E[\hat{\mu}_L] = \frac{1}{L} E[\sum_{i=1}^{L} X_i] = E[X]$ unbiased

- Variance :

$$\mathrm{var}(\hat{\mu}_L) = \mathrm{var}\left(\tfrac{1}{L}\textstyle\sum_{i=1}^{L} X_i\right)$$
$$= \frac{1}{L^2}\sum_{i=1}^{L}\mathrm{var}(X_i)$$
$$= \frac{\mathrm{var}(X_1)}{L} \sim \frac{1}{L}$$

decreases with sample size

- Standard deviation : $\mathrm{std}(\hat{\mu}_L) = \underset{=\sigma_L}{\sigma(\hat{\mu}_L)} = \sqrt{\mathrm{var}(\hat{\mu}_L)} \sim \frac{1}{\sqrt{L}}$



$P(|\hat{\mu}_L - \mu| > \varepsilon) \to 0$

$P(\hat{\mu}_L) \approx$ Gaussian for $L >> 1$

$P(\hat{\mu}_L \in [\mu - \sigma_L, \mu + \sigma_L]) \approx 0.68$

- Error bars : $\underset{\text{estimator}}{\hat{\mu}_L} \pm \underset{\text{error bar}}{\sigma_L}$ confidence interval at 68%

$\hat{\mu}_L \pm 2\sigma_L$ CI at 95%

- Estimator of σ_L :

$$\sigma_L = \frac{\sqrt{\mathrm{var}(X_1)}}{\sqrt{L}} \Rightarrow \hat{\sigma}_L = \frac{\hat{\sigma}_X}{\sqrt{L}}$$

$\sim \approx \frac{1}{L}$

$$\hat{\sigma}_X^2 = \frac{1}{L-1}\sum_{i=1}^{L}(X_i - \hat{\mu}_L)^2$$

$$\Rightarrow \hat{\sigma}_L = \underset{?}{\frac{1}{\sqrt{L}}} \sqrt{\underset{\overset{?}{\frac{1}{L-1}}}{\frac{1}{L}} \sum_{i=1}^{L} X_i^2 - \underbrace{\left(\frac{1}{L}\sum_{i=1}^{L} X_i\right)}_{\hat{\mu}_L}{}^2}$$