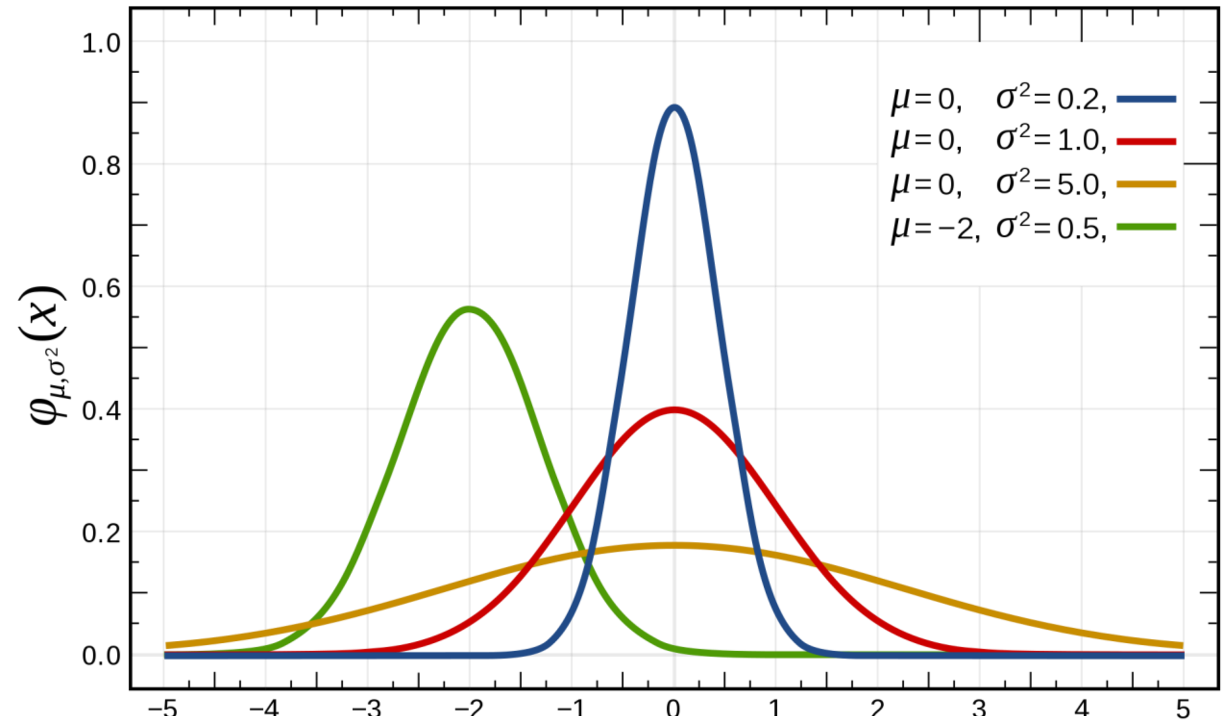
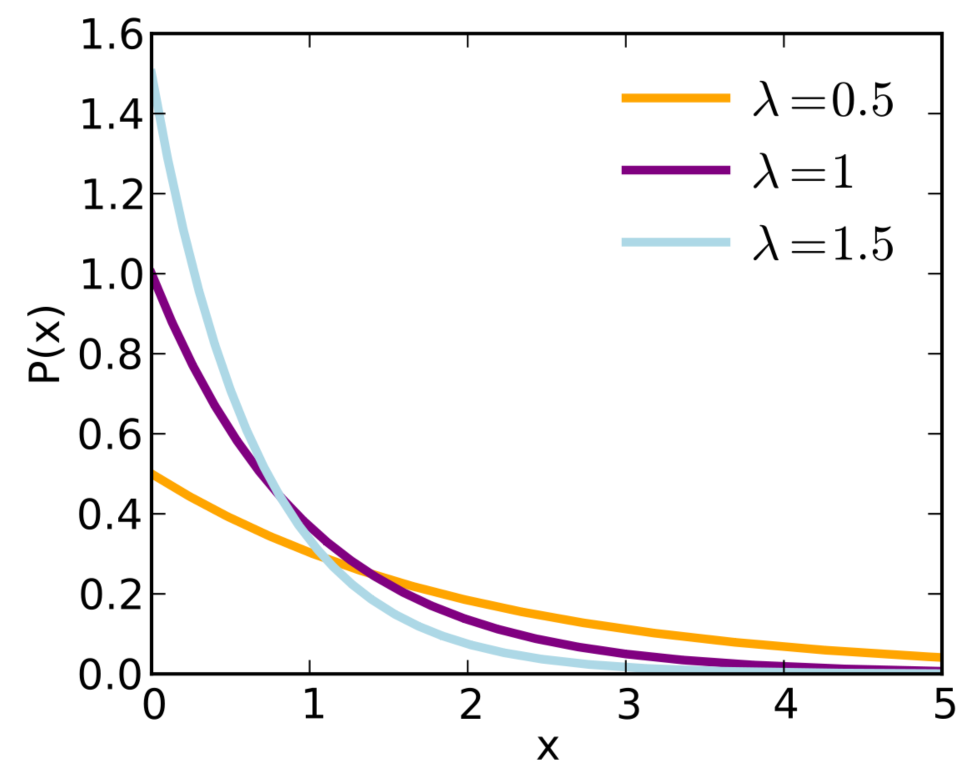
A Primer to Sampling and Estimation

Srijith P K

Sampling and Estimation

- What you have learnt : X is a random variable following a distribution
    - what does this mean ?
- Given the parameter value of a distribution how the density/distribution function look.
- Sampling : If X follows a distribution how can we obtain different value X will take
- Parameter estimation : Given different values X take, how can we obtain the parameters of the underlying distribution

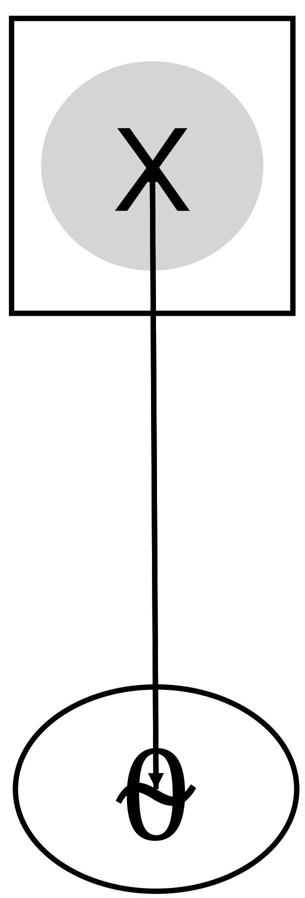


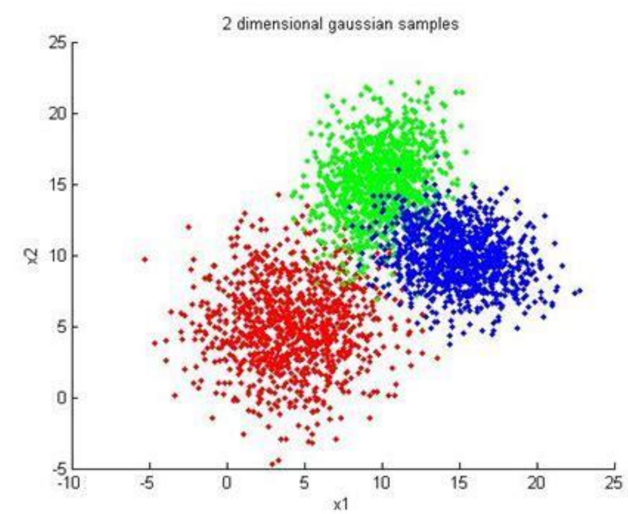


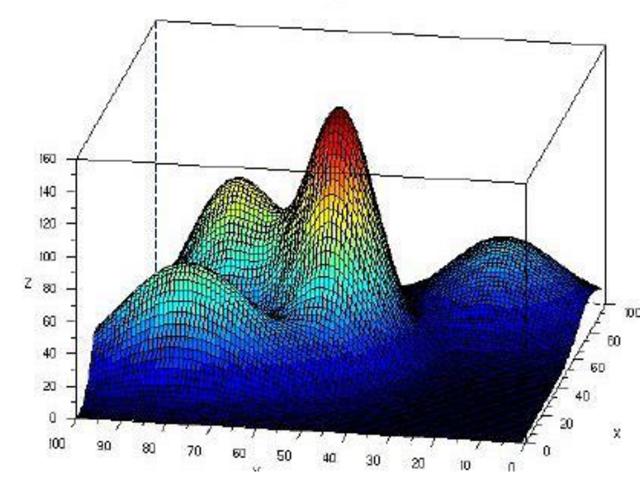
Sampling and Estimation

- Given data Statistics/ML aim to learn parameters of the underlying distribution from data
  - Gaussian mixture model, probabilistic graphical models, linear regression, logistic regression etc.
- Sampling is essential in probabilistic machine learning and Bayesian statistics
  - Latent dirichlet allocation, Gaussian process, probabilistic graphical model
  - Also used in computational physics, biology and many engineering disciplines.

    MCMC was placed in the top 10 most important algorithms of the 20th century

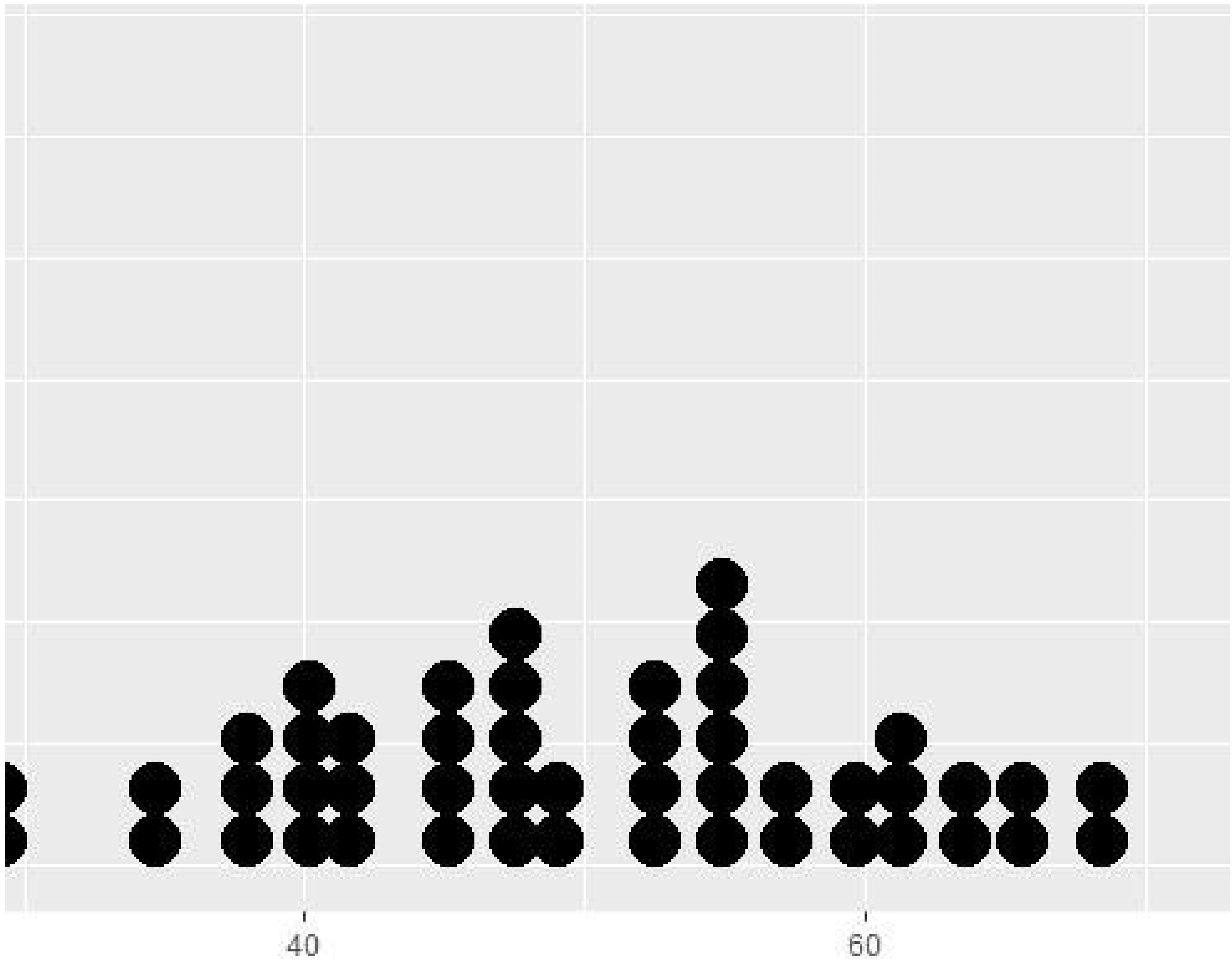






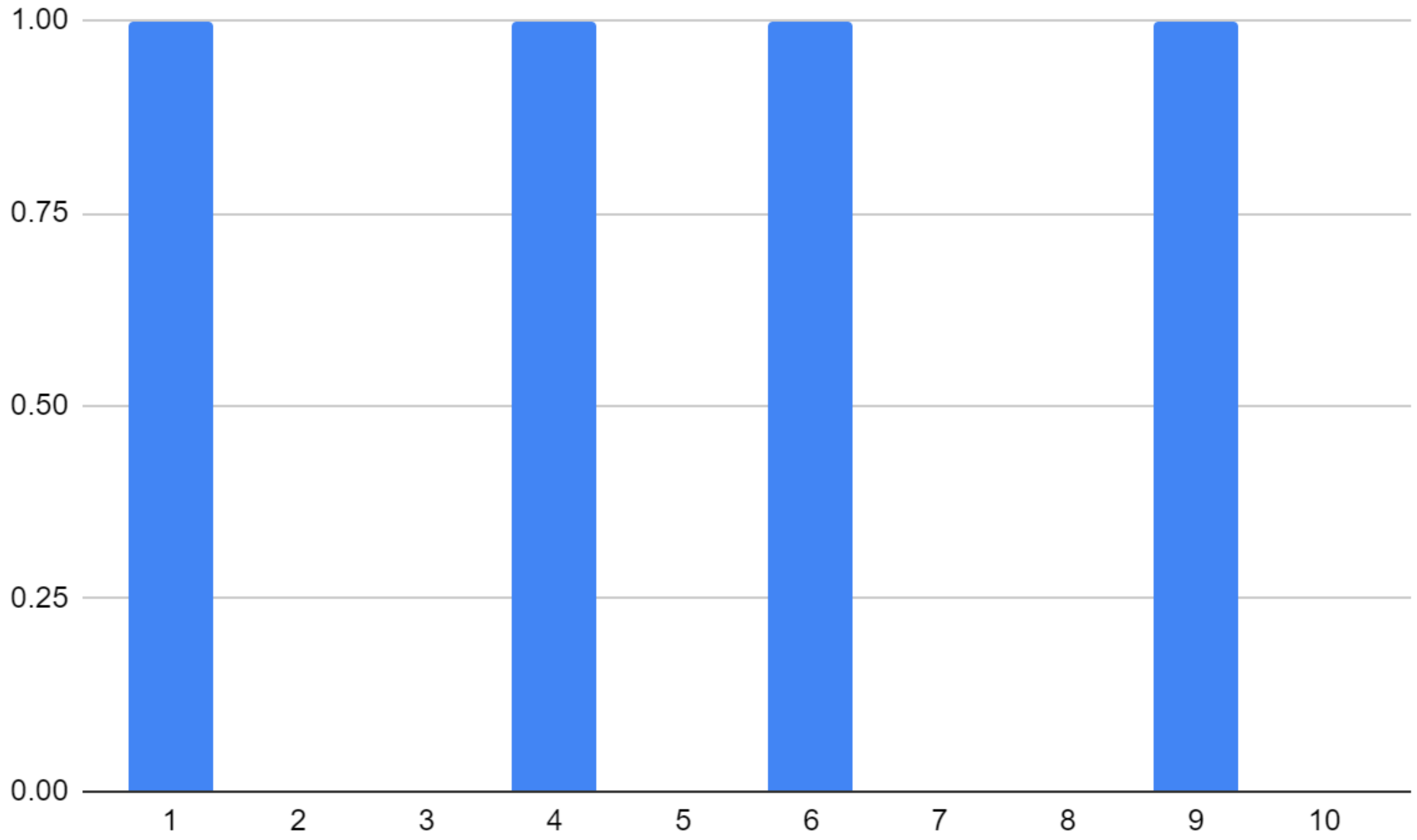
Parameter Estimation

- Data points representing the weight (in kgs) of students in a class.
- Whats mean and std deviation of the data ?
- Whats the probability that weight > 60



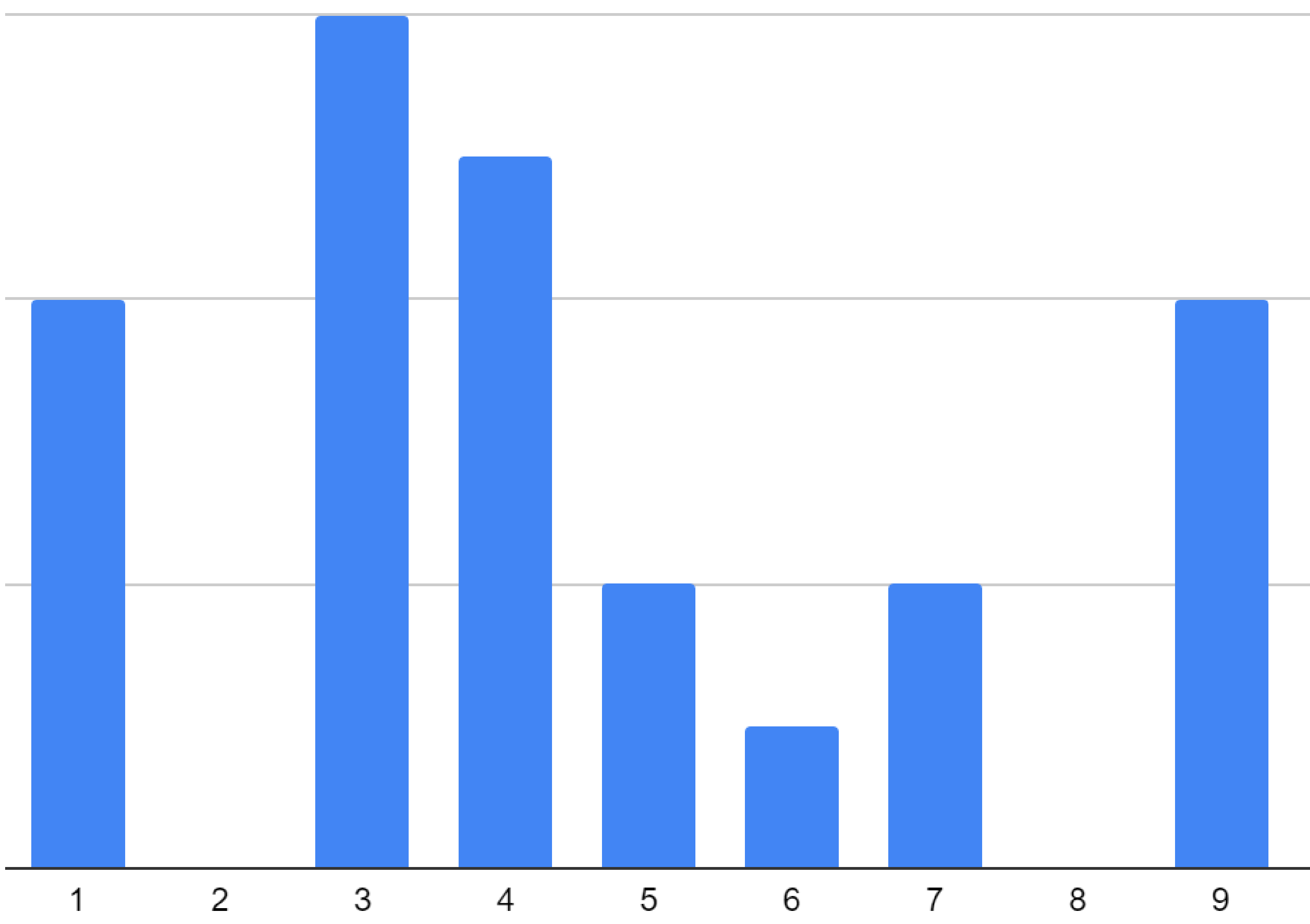
Parameter Estimation

- Data points representing the if it has rained or not in last 10 days.
- Whats the probability that it will rain tomorrow ?
- How many days will it rain in next 5 days ?



Parameter Estimation

- The number of traffic accidents in Berkeley, California, in 10 randomly chosen nonrainy days in 1998 is as follows:
- 4, 0, 6, 5, 2, 1, 2, 0, 4, 3
- Use these data to estimate the proportion of nonrainy days that had 2 or fewer accidents that year.



Parameter Estimation

- Any statistic used to estimate the value of an unknown parameter θ is called an estimator of θ.
    - mean and variance for Normal, rate (lambda) for Poisson, etc.

Parameter Estimation

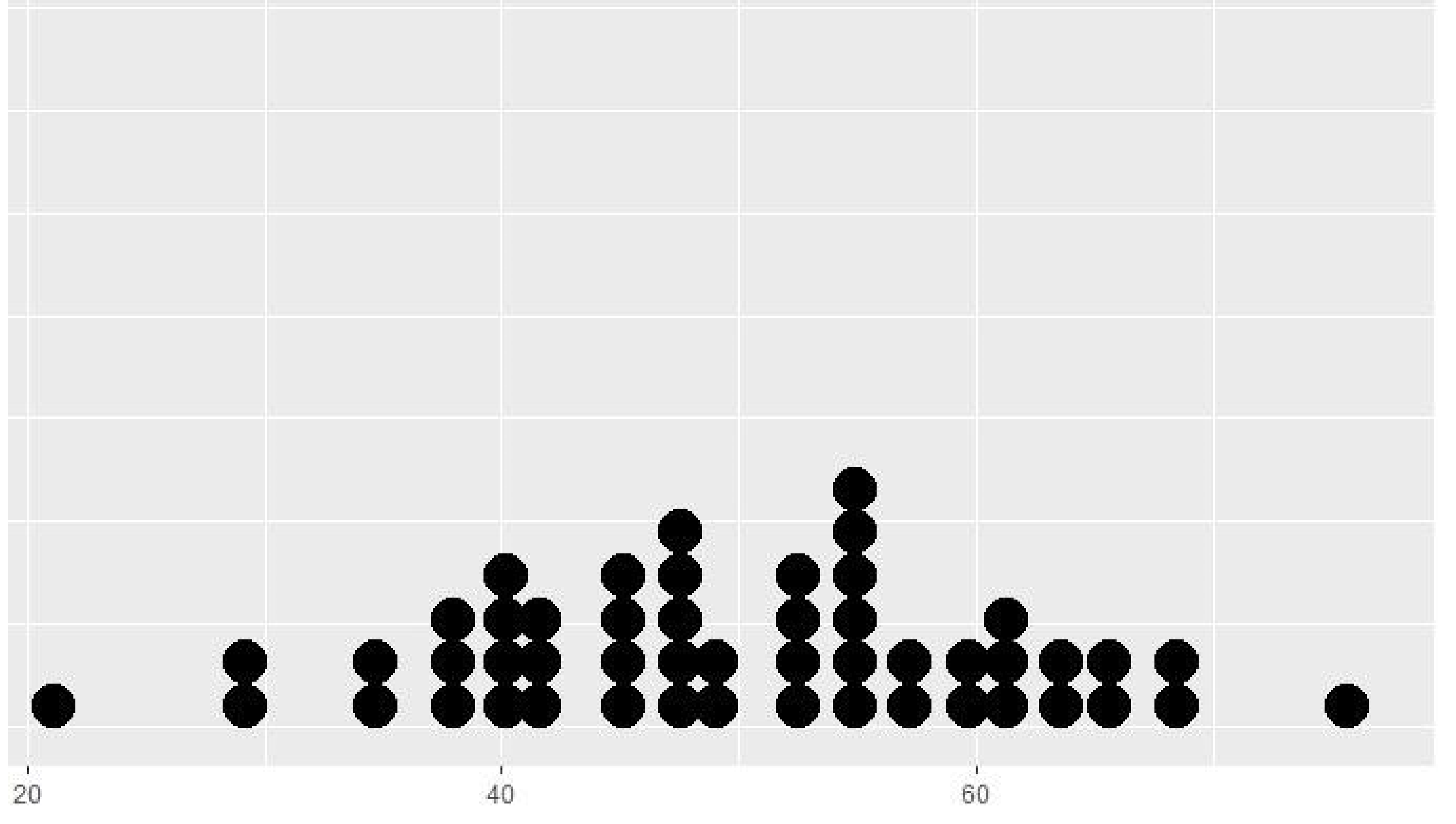
- Any statistic used to estimate the value of an unknown parameter θ is called an estimator of θ.
    - mean and variance for Normal, rate (lambda) for Poisson, etc.
- **Maximum likelihood** estimator
- MLE can be defined as a method for estimating population parameters from sample data such that the probability (likelihood) of obtaining the observed data is maximized.

Parameter Estimation

- Any statistic used to estimate the value of an unknown parameter θ is called an estimator of θ.
  - mean and variance for Normal, rate (lambda) for Poisson, etc.
- **Maximum likelihood** estimator
- MLE can be defined as a method for estimating population parameters from sample data such that the probability (likelihood) of obtaining the observed data is maximized.
- f (x1, . . . , xn|θ) represents the probability or likelihood that the values x1, x2, . . . , xn will be observed when θ is the true value of the parameter, maximum likelihood estimate ˆθ is defined to be that value of θ maximizing f (x1, . . . , xn|θ)
- Maximum Likelihood estimation :
- argmax_θ p(x|θ) = argmax_θ log p(x|θ)

Parameter estimation

- which of the following would maximize the probability of observing the data
  - Mean = 100, SD = 10
  - Mean = 50, SD = 10



Maximum likelihood estimation

- (Maximum Likelihood Estimator of a Bernoulli Parameter) Suppose you have data from n independent Bernoulli trials, $X1, \ldots, Xn$. Assuming the success probability is p what is the maximum likelihood estimator of p?

$$X_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{otherwise} \end{cases}$$

$$P\{X_i = 1\} = p = 1 - P\{X_i = 0\}$$

$$P\{X_i = x\} = p^x(1-p)^{1-x}, \quad x = 0, 1$$

$$\begin{aligned} f(x_1, \ldots, x_n|p) &= P\{X_1 = x_1, \ldots, X_n = x_n|p\} \\ &= p^{x_1}(1-p)^{1-x_1} \cdots p^{x_n}(1-p)^{1-x_n} \\ &= p^{\Sigma_1^n x_i}(1-p)^{n-\Sigma_1^n x_i}, \quad x_i = 0, 1, \quad i = 1, \ldots, n \end{aligned}$$

Maximum likelihood estimation

- (Maximum Likelihood Estimator of a Bernoulli Parameter) Suppose you have data from n independent Bernoulli trials, $X1, \ldots, Xn$. Assuming the success probability is p what is the maximum likelihood estimator of p?

$$X_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{otherwise} \end{cases} \qquad P\{X_i = 1\} = p = 1 - P\{X_i = 0\}$$

To determine the value of p that maximizes the likelihood,

$$\log f(x_1, \ldots, x_n|p) = \sum_1^n x_i \log p + \left(n - \sum_1^n x_i\right) \log(1-p)$$

Maximum likelihood estimation

- (Maximum Likelihood Estimator of a Bernoulli Parameter) Suppose you have data from n independent Bernoulli trials, *X*1, . . . , *Xn*. Assuming the success probability is p what is the maximum likelihood estimator of p?

$$X_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{otherwise} \end{cases}$$

$$P\{X_i = 1\} = p = 1 - P\{X_i = 0\}$$

To determine the value of p that maximizes the likelihood,

$$\log f(x_1, \ldots, x_n | p) = \sum_1^n x_i \log p + \left(n - \sum_1^n x_i\right) \log(1 - p)$$

$$\frac{d}{dp} \log f(x_1, \ldots, x_n | p) = \frac{\sum_1^n x_i}{p} - \frac{\left(n - \sum_1^n x_i\right)}{1 - p}$$

$$\hat{p} = \frac{\sum_{i=1}^n x_i}{n}$$

Maximum likelihood estimation

- (Maximum Likelihood Estimator of a Bernoulli Parameter) Suppose you have data from n independent Bernoulli trials, *X*1, . . . , *Xn*. Assuming the success probability is p what is the maximum likelihood estimator of p?

$$X_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{otherwise} \end{cases}$$

$$P\{X_i = 1\} = p = 1 - P\{X_i = 0\}$$

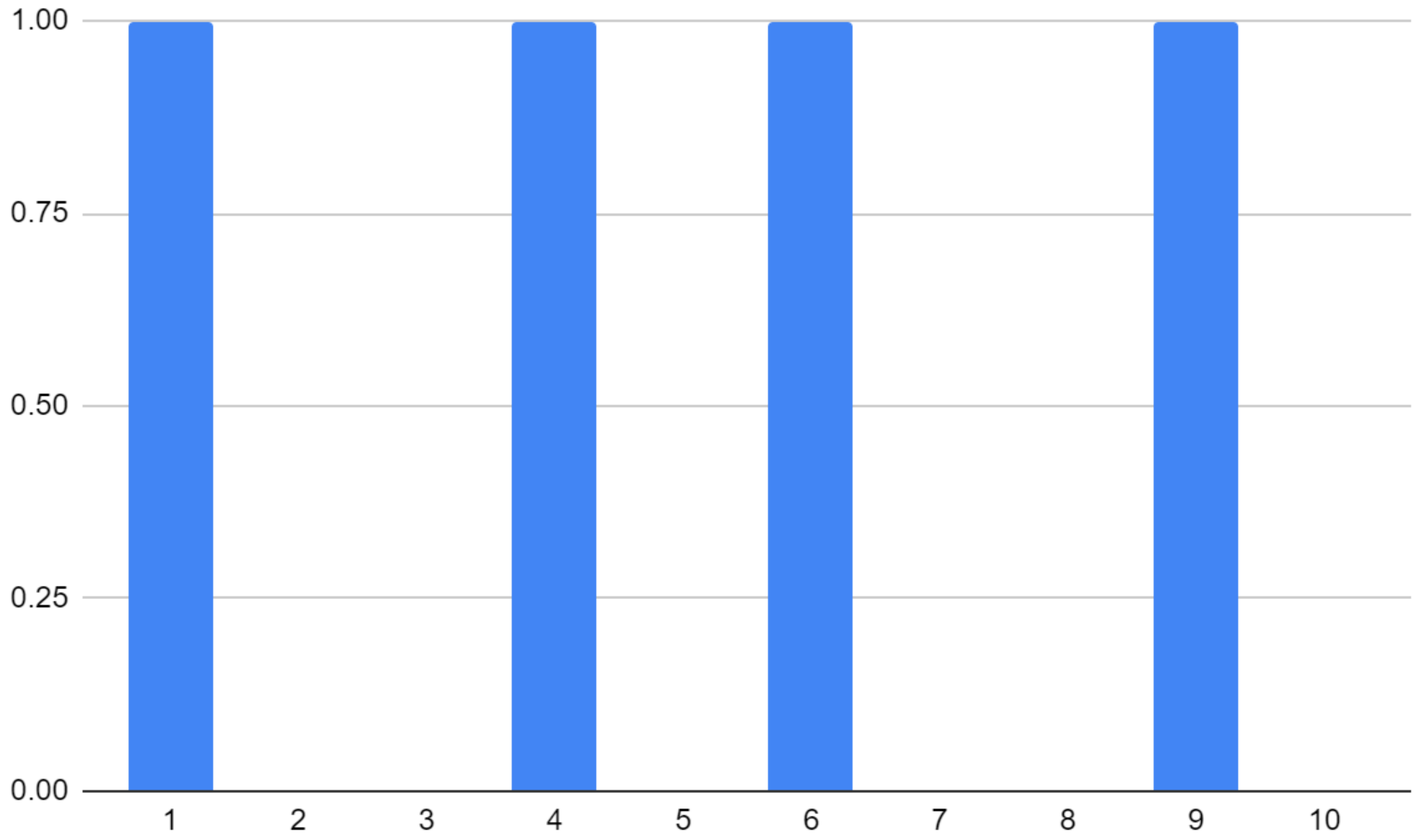
To determine the value of p that maximizes the likelihood,

proportion of the observed trials that result in successes.

$$\frac{d}{dp} \log f(x_1, \ldots, x_n | p) = \frac{\sum_1^n x_i}{p} - \frac{\left(n - \sum_1^n x_i\right)}{1-p}$$

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

- Data points representing the if it has rained or not in last 10 days.
- Whats the probability that it will rain tomorrow ?
- How many days will it rain in next 5 days ?



Parameter Estimation

- Multinomial
- 3,1,2,4,3,5,6,1,3,4

Maximum likelihood estimation

- (Maximum Likelihood Estimator of a Poisson Parameter) Suppose $X1, \ldots, Xn$ are independent Poisson random variables each having mean λ. Determine the maximum likelihood estimator of λ.

Maximum likelihood estimation

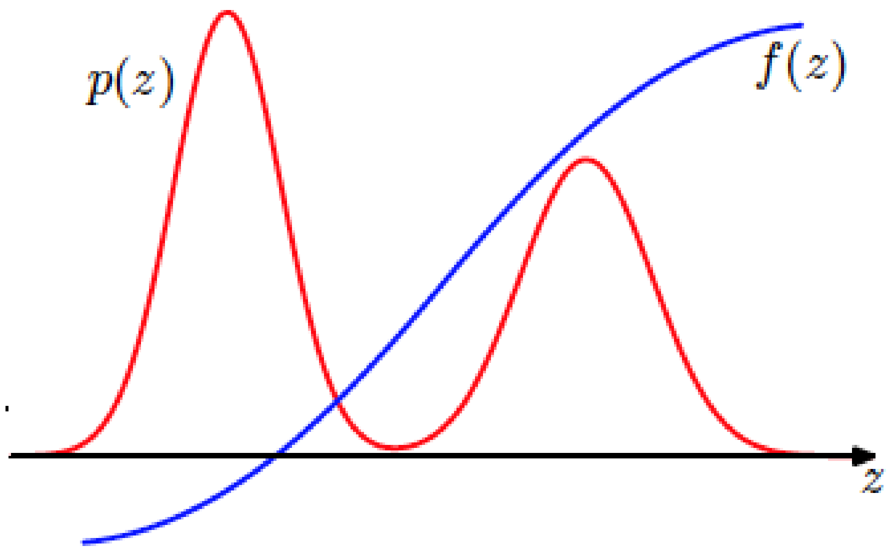
- The number of traffic accidents in Berkeley, California, in 10 randomly chosen nonrainy days in 1998 is as follows:
- 4, 0, 6, 5, 2, 1, 2, 0, 4, 3
- Use these data to estimate the proportion of nonrainy days that had 2 or fewer accidents that year.

Maximum likelihood estimation

- (Maximum Likelihood Estimator in a Normal Population) Suppose $X_1, \ldots, X_n$ are independent, normal random variables each with unknown mean μ and unknown standard deviation σ.

Sampling

- Monte carlo approximation for expectation and optimization – named after a city in france
- z^l are samples generated from p(z)
- first developed in the area of statistical physics - during development of the atomic bomb
- Find marginal distribution from joint, expectatio random variables,, expectation of function of random variables, transformation of random variables……
- Optimization of complicated functions

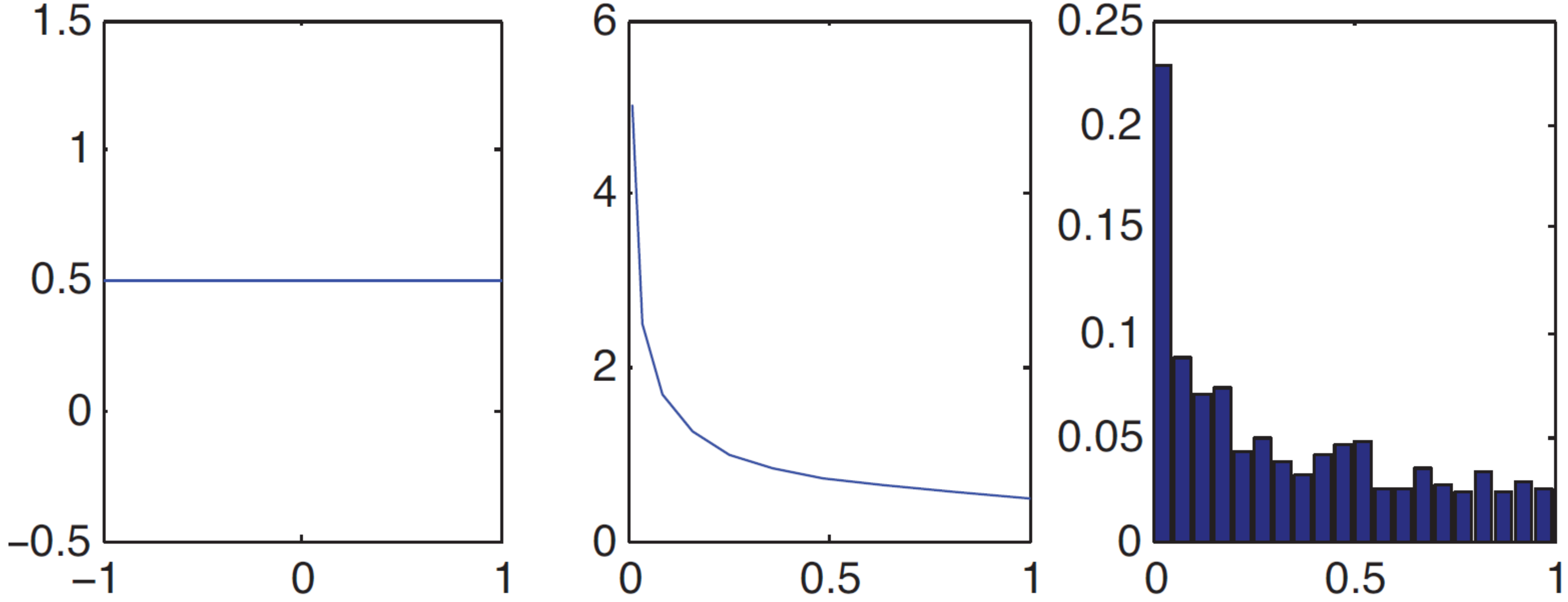


$$\mathbb{E}[f] = \int f(\mathbf{z})p(\mathbf{z})\,\mathrm{d}\mathbf{z}$$

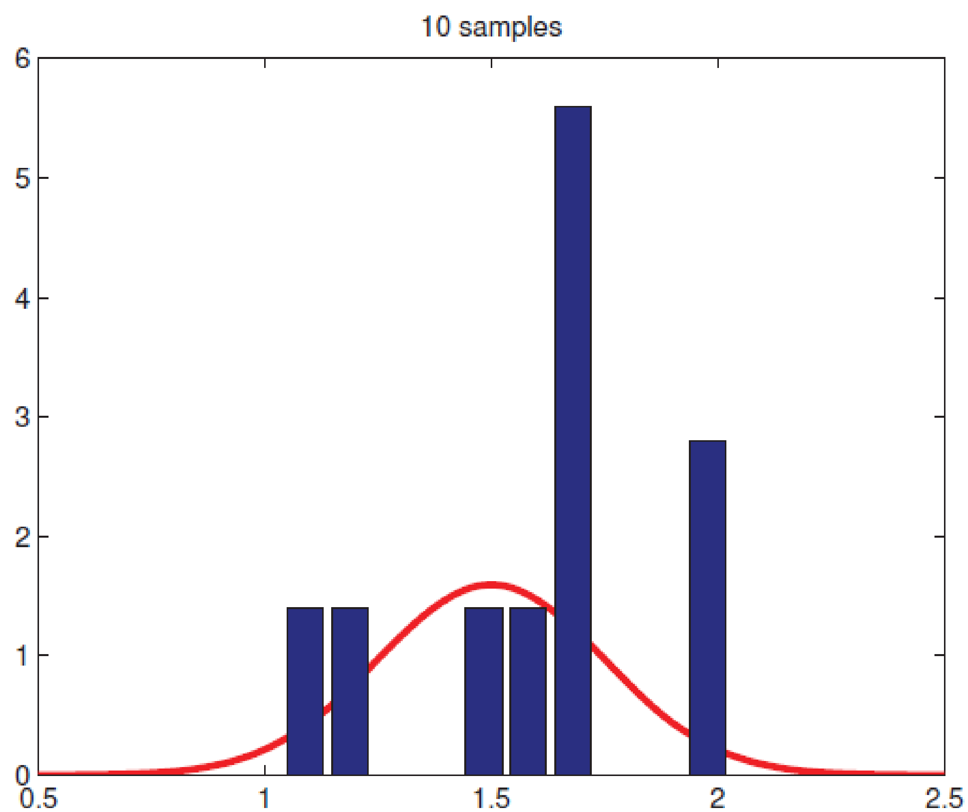
$$\widehat{f} = \frac{1}{L}\sum_{l=1}^{L} f(\mathbf{z}^{(l)}).$$

Sampling

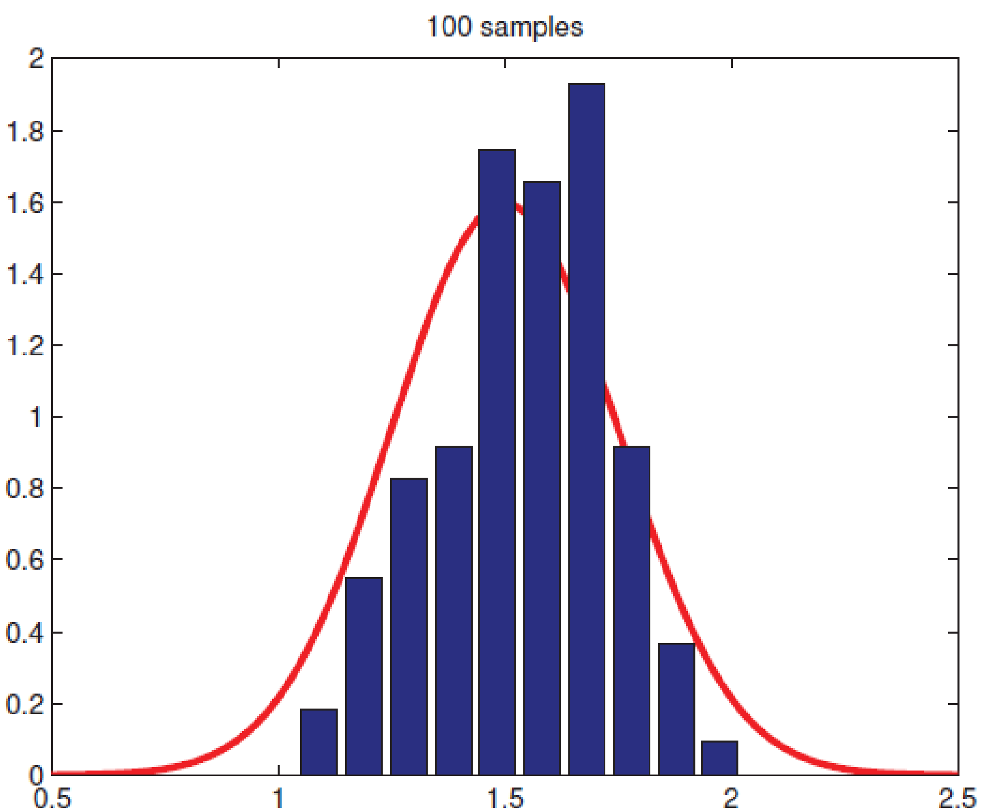
- Transformation of random variable $y = x^2,$



Sampling



(a)



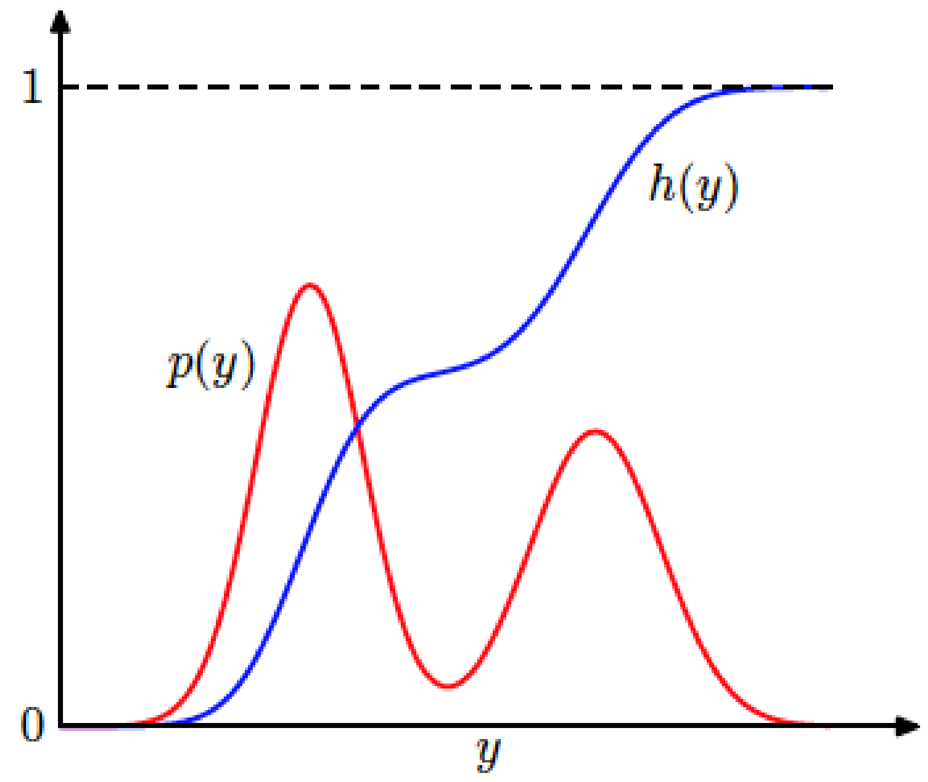
(b)

$$\mathcal{N}(\mu = 1.5, \sigma^2 = 0.25).$$

how do we efficiently generate samples from a probability distribution ?

Inverse Probabiity Transform

Uses cumulative distribution function (CDF

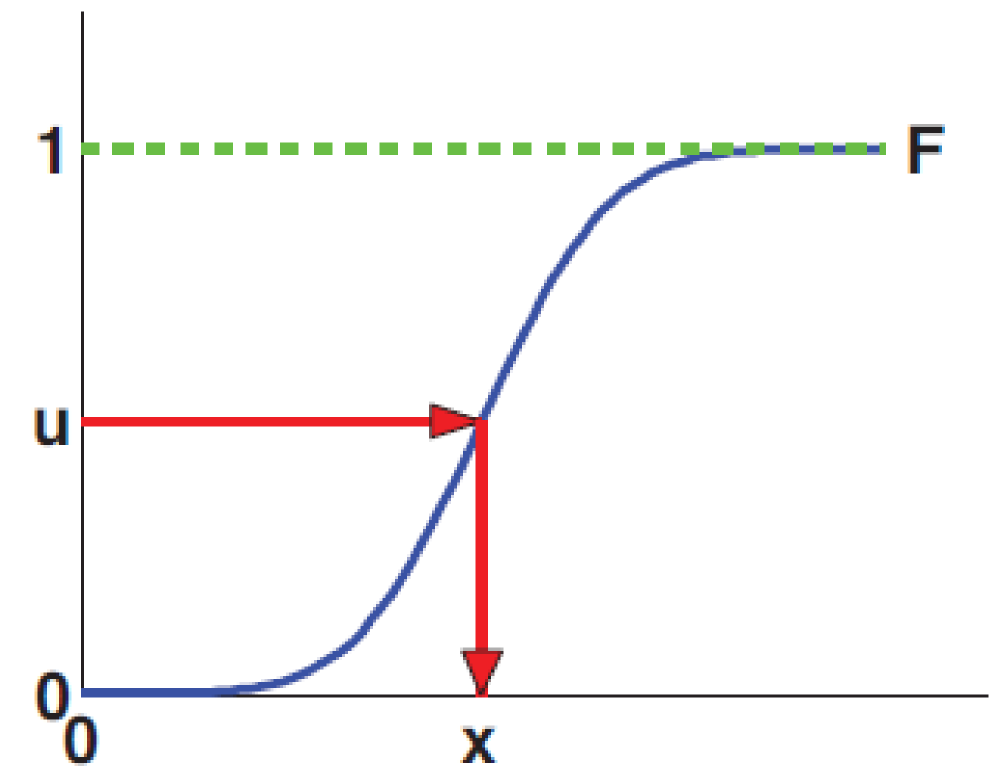


Inverse Probabiity Transform

Uses cumulative distribution function (CDF)

$$\begin{aligned}\Pr(F^{-1}(U) \le x) &= \Pr(U \le F(x)) \\ &= F(x)\end{aligned}$$

generate a random number $u \sim U(0, 1)$
using a **pseudo random number Generator**

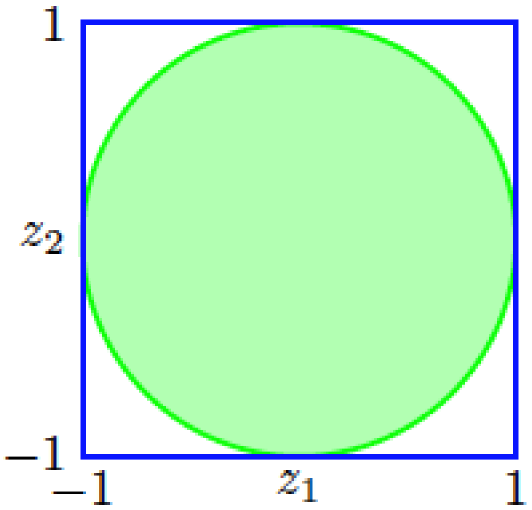


Sampling : Exponential

- Obtain samples from the exponential distribution $p(y) = \lambda \exp(-\lambda y)$

Sampling : Gaussian

- Box-Muller method : N(y ; \mu, \Sigma)
- generate pairs of uniformly distributed random numbers z1, z2 ∈ (−1, 1)
- Obtain uniform distribution of points inside the unit circle with p(z1, z2) = 1/π



Sampling : Gaussian

- Consider transformation

$$
\begin{aligned}
y_1 &= z_1 \left(\frac{-2 \ln z_1}{r^2} \right)^{1/2} \\
y_2 &= z_2 \left(\frac{-2 \ln z_2}{r^2} \right)^{1/2}
\end{aligned}
$$

$$r^2 = z_1^2 + z_2^2.$$

$$
\begin{aligned}
p(y_1, y_2) &= p(z_1, z_2) \left| \frac{\partial(z_1, z_2)}{\partial(y_1, y_2)} \right| \\
&= \left[\frac{1}{\sqrt{2\pi}} \exp(-y_1^2/2) \right] \left[\frac{1}{\sqrt{2\pi}} \exp(-y_2^2/2) \right]
\end{aligned}
$$

$$\mathbf{y} = \boldsymbol{\mu} + \hat{\mathbf{L}}\mathbf{z} \qquad \boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$$


THANK YOU