Poisson approximation

1 Poisson approximation

Poisson distribution $Po(\lambda)$

PDF and CDF of Poisson distribution $Po(\lambda)$

Poisson approximation

Example - Number of couples with same birthday

2 Construction of Poisson point process using a p-coin

Construction of Poisson point process $PPP(\lambda)$ with intensity λ using a p-coin

Definition of Poisson point process $PPP(\lambda)$ with intensity λ

3 Discrete vs continuous random variables

Discrete vs continuous random variables - Distribution

Discrete vs continuous random variables - Joint distribution

Discrete vs continuous random variables - Joint, marginal, conditional

Discrete vs continuous random variables - Independence

4 CDF and quantile

CDF and quantile

Quartile Q_i

Box plot

Example - CDF of $U(0,1)$

VaR and CVaR

5 Exponential approximation

Exponential distribution $Exp(\lambda)$

Exponential approximation

Memoryless property of geometric and exponential distribution

6 Construction of Poisson point process using an $Exp(\lambda)$-coin

Construction of Poisson point process $PPP(\lambda)$ using an $Exp(\lambda)$-coin

Example - Joint PDF of independent random variables

Example - Hike over Mt. Bukhan

7 Thinning and merger

Poisson distribution $Po(\lambda)$

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{for } k = 0, 1, 2, \ldots$$

Distribution	Random variable
$B(p)$	Flip a p-coin and check whether we have a head
$B(n,p)$	Flip a p-coin n times and count the number of heads
$Po(\lambda) \approx B(n,p)$	Flip a p-coin n times and count the number of heads, where $np = \lambda$ is fixed and $n \to \infty$
$Geo(p)$	Flip a p-coin until first head and count the number of flips
$NB(r,p)$	Flip a p-coin until r-th head and count the number of flips

Distribution	Expectation	Variance
$B(p)$	p	pq
$B(n,p)$	np	npq
$Po(\lambda) \approx B(n,p)$	λ	λ
$Geo(p)$	$\frac{1}{p}$	$\frac{q}{p^2}$
$NB(r,p)$	$\frac{r}{p}$	$\frac{rq}{p^2}$

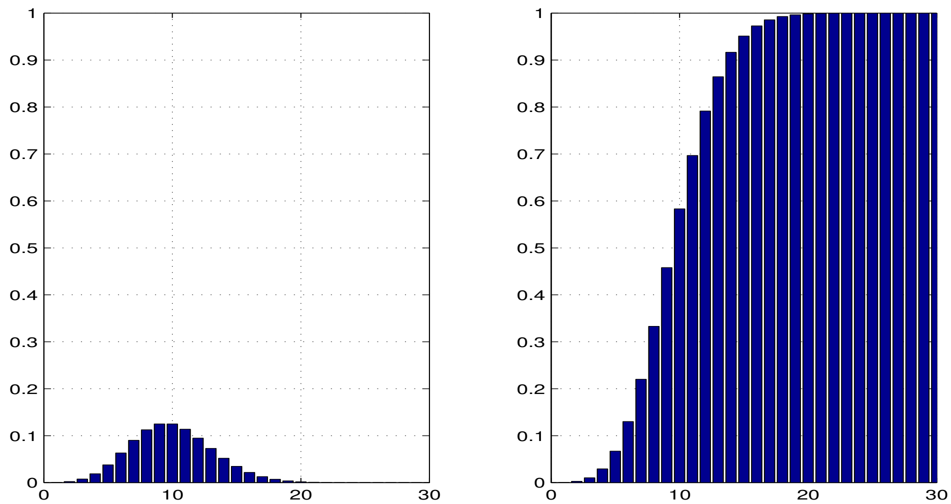


Figure 1: PDF (left) and CDF (right) of $Po(10)$

```
clear all; close all; clc;

la=10;

% Maximum range 0:m considered for Poisson distribution
m=30;

% PMF for Poisson distribution
x=0:m;
PMF=exp(-la).*(la.^x)./factorial(x);
CDF=cumsum(PMF);

subplot(1,2,1) % PMF
bar(x,PMF); axis([0 m 0 1]); grid on;

subplot(1,2,2) % CDF
bar(x,CDF); axis([0 m 0 1]); grid on;
```

Poisson approximation

	Random variable
$B(n,p)$	Flip a p-coin n times and count the number of heads
$Po(\lambda){\approx}B(n,p)$	Flip a p-coin n times and count the number of heads, where n is large, p is small, $np = \lambda$ is medium

	Expectation	Variance
$B(n,p)$	np	npq
$Po(\lambda){\approx}B(n,p)$	λ	λ

Error bound

A_i	Independent events with $p_i = P(A_i)$
1_{A_i}	Independent random variables with $1_{A_i} \sim B(p_i)$
$X = \sum_{i=1}^{n} 1_{A_i}$	not $B(n,p)$ in general

Then, with $Y \sim Po(\lambda)$, $\lambda = \sum_{i=1}^{n} p_i$, for any A

$$|P(X \in A) - P(Y \in A)| \le \sum_{i=1}^{n} p_i^2 \le \left(\max_{1\le i\le n} p_i\right)\left(\sum_{i=1}^{n} p_i\right) = \left(\max_{1\le i\le n} p_i\right)\lambda$$

$$\begin{aligned}
P(X=k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \frac{1}{k!}\frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\lambda^k \left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-k} \\
&\to \frac{1}{k!}\cdot 1^k \cdot \lambda^k \cdot e^{-\lambda}\cdot 1 = \frac{\lambda^k}{k!}e^{-\lambda}
\end{aligned}$$

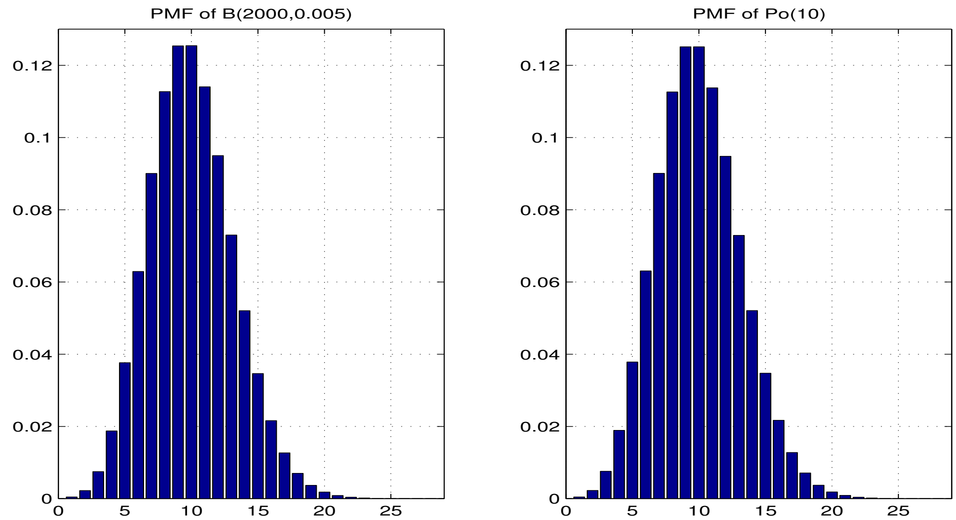


Figure 2: Poisson approximation of $B(n, p)$, $n = 2,000$, $p = 0.005$, by $Po(\lambda)$, $\lambda = 10$. Maximum difference between these two PMF is 3.1393×10^{-4}.

```
clear all; close all; clc;

n=2000; p=0.005; q=1-p; la=n*p;

% Maximum range 0:m considered for both Binomial and Poisson distribution
m=ceil(la+6*sqrt(la));

subplot(1,2,1) % PMF of B(n,p)
i=0:m;
PMF_B=binomial(n,i).*(p.^i).*(q.^(n-i));
bar(0:m,PMF_B); axis([0 m 0 0.13]); grid on;
title('PMF of B(2000,0.005)')

subplot(1,2,2) % PMF of B(n,p)
i=0:m;
PMF_P=exp(-la)*(la.^i)./factorial(i);
bar(0:m,PMF_P); axis([0 m 0 0.13]); grid on;
title('PMF of B(2000,0.005)')

% Difference between PMF of B(n,p) and Po(la)
PMF_Diff=PMF_B-PMF_P;
Max_PMF_Diff=max(abs(PMF_Diff))
```

Example - Number of couples with same birthday

Approximately 80,000 marriage took place in NY. Estimate the probability that there are more than 250 couples with same birthday who married in NY last year.

With $n = 80,000$, $p = 1/365$, $\lambda = np$, Let S_n be the number of couples with same birthday. Let A_i be the event that the ith couple share the common birthday and let 1_{A_i} be its indicator. Then,

$$S_n = \sum_{i=1}^{n} 1_{A_i} \sim B(n,p) \approx Po(\lambda)$$

Exact probability using $S_n \sim B(n,p)$ $\qquad P(S_n > 250) = 0.0187$

Approximate probability using $X \sim Po(\lambda)$ $\qquad P(X > 250) = 0.0188$

```
clear all; close all; clc;

n=80000; p=1/365; q=1-p; la=n*p;
m=250;

tic
Prob=q^n;
Cum_Prob=Prob;
for i=1:m
    Prob=Prob*(n-(i-1))/i*p/q;
    Cum_Prob=Cum_Prob+Prob;
end
Binomial_Exact_Prob=1-Cum_Prob
Binomial_Exact_Prob_Computing_Time=toc

tic
Prob=exp(-la);
Cum_Prob=Prob;
for i=1:m
    Prob=Prob*la/i;
    Cum_Prob=Cum_Prob+Prob;
end
Poisson_Approximate_Prob=1-Cum_Prob
Poisson_Approximate_Prob_Computing_Time=toc
```

Construction of Poisson point process $PPP(\lambda)$ with intensity λ using a p-coin

Number of ticks per year	n
p-coin flip at tick k	H with marker
Number of ticks between 0 and t	nt
Number $N([0,t])$ of markers between 0 and t	$B(nt,p) \approx Po(\lambda t)$
Number of ticks between s and t	$n(t-s)$
Number $N([s,t])$ of markers between s and t	$B(n(t-s),p) \approx Po(\lambda(t-s))$

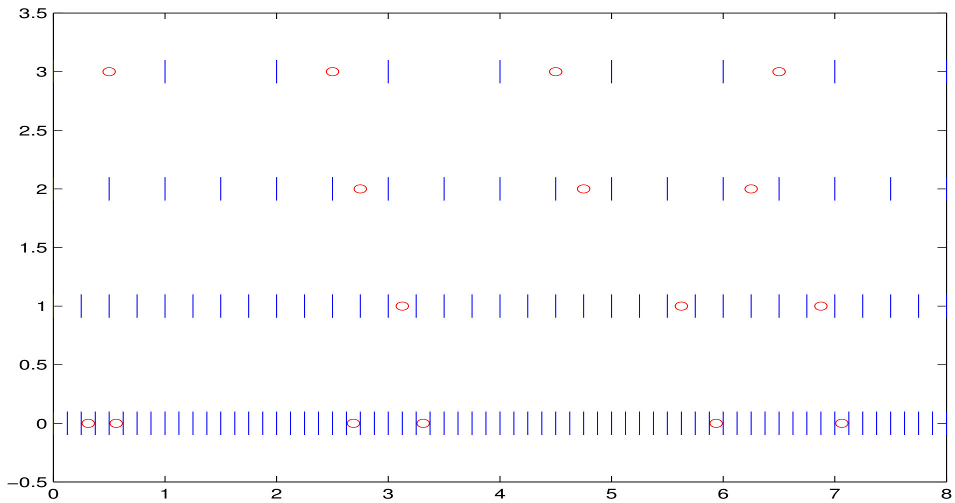


Figure 3: Construction of Poisson point process

```
clear all; close all; clc; rng('default')

la=0.5;

% Maximum range 0:n considered for Poisson point process
n=8;

% Poisson point process layer
m=4;

for i=1:m

    x=rand(n*2^(i-1),1);
    p=la/2^(i-1);
    index=find(x<p);
    position=(index-0.5)/2^(i-1);
    plot(position,(m-i)*ones(length(position),1),'or'); hold on;

    for j=0:2^(-i+1):n
        line([j j],(m-i)+[-0.1 0.1]); hold on;
    end

end
```

Definition of Poisson point process $PPP(\lambda)$ with intensity λ

For each finite interval I we attach a random variable $N(I)$. A collection of random variables $N(I)$ is the Poisson point process $PPP(\lambda)$ with intensity λ if

(1) $N([s,t]) \sim Po(\lambda(t-s))$

(2) For any $t_0 < t_1 < t_2 < \cdots < t_m$, $N([t_{i-1}, t_i])$ are all independent

$\boxed{N([s,t]) \sim Po(\lambda(t-s))}$

$$N([s,t]) \sim B(n(t-s), p) \approx Po(\lambda(t-s)) \quad \text{if } n \text{ goes to the infinite}$$

$\boxed{N([0,t]) \text{ has independent increments}}$

For any $t_0 < t_1 < t_2 < \cdots < t_m$ the coin flips in one time interval $[t_i, t_{i-1}]$ are completely different from the coin flips in other time interval $[t_j, t_{j-1}]$. So,

$$N([t_{i-1}, t_i]) \text{ are all independent}$$

Discrete vs continuous random variables - Distribution

$$\sum \text{ with discrete random variable} \quad \Leftrightarrow \quad \int \text{ with continuous random variable}$$

	Discrete random variable	Continuous random variable
Possible values	Discrete	Continuous
PMF/PDF	$0 \leq p_x \leq 1$	$0 \leq f(x) \leq \infty$

Meaning

$$\mathbb{P}(X = x) = p_x, \qquad \mathbb{P}(x \leq X \leq x + dx) = f(x)dx$$

Total mass 1

$$\sum_x p_x = 1, \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

Expectation

$$\mathbb{E}X = \sum_x x p_x, \qquad \mathbb{E}X = \int_{-\infty}^{\infty} x f(x)dx$$

k-th moment

$$\mathbb{E}X^k = \sum_x x^k p_x, \qquad \mathbb{E}X^k = \int_{-\infty}^{\infty} x^k f(x)dx$$

Variance

$$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2, \qquad Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

CDF

$$\mathbb{P}(X \leq x) = \sum_{s \leq x} p_s, \qquad \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(s)ds$$

Discrete vs continuous random variables - Joint distribution

$\sum$ with discrete random variable $\Leftrightarrow$ $\int$ with continuous random variable

	Discrete random variable	Continuous random variable
Possible values	Discrete	Continuous
Joint	$0 \le p_{x,y} \le 1$	$0 \le f_{X,Y}(x,y) \le \infty$

Meaning

$\mathbb{P}(X = x, Y = y) = p_{x,y}$, $\qquad \mathbb{P}(x \le X \le x+dx, y \le Y \le y+dy) = f_{X,Y}(x,y)dxdy$

Total mass 1

$$\sum_x \sum_y p_{x,y} = 1, \qquad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$$

Expectation

$$\mathbb{E}X = \sum_x \sum_y x p_{x,y}, \qquad \mathbb{E}X = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy$$

k-th moment

$$\mathbb{E}X^k = \sum_x \sum_y x^k p_{x,y}, \qquad \mathbb{E}X^k = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k f_{X,Y}(x,y)dxdy$$

Variance

$$Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2, \qquad Var(X) = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

CDF

$$\mathbb{P}(X \le x) = \mathbb{P}(X \le x) = \sum_{s \le x} \sum_y p_{s,y}, \qquad \mathbb{P}(X \le x) = \int_{-\infty}^{\infty}\int_{-\infty}^{x} f(s,y)dsdy$$

Discrete vs continuous random variables - Joint, marginal, conditional

$$\sum \text{ with discrete random variable} \quad \Leftrightarrow \quad \int \text{ with continuous random variable}$$

$\boxed{\text{From joint to marginal}}$

$$p_{x_i} = \sum_{y_j} p_{x_i,y_j}, \qquad f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

$$p_{y_j} = \sum_{x_i} p_{x_i,y_j}, \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

$\boxed{\text{From joint to conditional}}$

$$p_{x_i|y_j} = \frac{p_{x_i,y_j}}{\sum_{x_{i'}} p_{x_{i'},y_j}}, \qquad f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$p_{y_j|x_i} = \frac{p_{x_i,y_j}}{\sum_{y_{j'}} p_{x_i,y_{j'}}}, \qquad f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

$\boxed{\text{From marginal and conditional to joint}}$

$$p_{x_i,y_j} = p_{x_i} p_{y_j|x_i}, \qquad f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y|x)$$

$$p_{x_i,y_j} = p_{y_j} p_{x_i|y_j}, \qquad f_{X,Y}(x,y) = f_Y(y) f_{X|Y}(x|y)$$

Discrete vs continuous random variables - Independence

$$\sum \text{ with discrete random variable} \quad \Leftrightarrow \quad \int \text{ with continuous random variable}$$

How to check independency

$$\text{For PMF} \qquad p_{x_i,y_j} = p_{x_i} p_{y_j}$$

$$\text{For PDF} \qquad f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

CDF and quantile

F	CDF
F^{-1}	Quantile function

CDF

A CDF F has the following three properties:

$$\begin{array}{ll} (1) & x_1 < x_2 \quad \Rightarrow \quad F(x_1) \le F(x_2) \\ (2) & \lim_{x \to -\infty} F(x) = 0 \;\text{ and }\; \lim_{x \to \infty} F(x) = 1 \\ (3) & \lim_{x \downarrow x_0} F(x) = F(x_0) \;\text{ and }\; \lim_{x \uparrow \infty} F(x) = F(x_0-) \end{array}$$

Conversely, if a function F satisfies the above three, then we can interpret F as a CDF of a distribution. If F is a CDF, then it satisfies

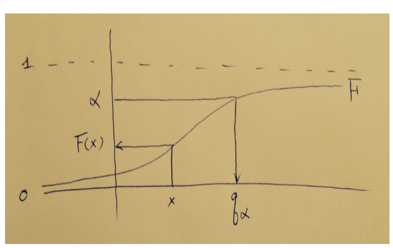
$$\begin{array}{ll} (4) & F(x_0) - F(x_0-) = \mathbb{P}(X = x_0) \\ (5) & \frac{d}{dx} F(x) = f(x) \quad \text{if } X \text{ has the PDF } f \\ (6) & P(a \le X \le b) = F(b) - F(a) \quad \text{if } X \text{ is continuous} \\ (7) & P(X \ge x) = 1 - P(X \le x) = 1 - F(x) \quad \text{if } X \text{ is continuous} \end{array}$$

Quantile q_α

The α quantile of F is the value q_α such that

$$F(q_\alpha) = \mathbb{P}(X \le q_\alpha) = \alpha$$

1% quantile $q_{0.01}$	$q_{0.01} = F^{-1}(0.01)$
5% quantile $q_{0.05}$	$q_{0.05} = F^{-1}(0.05)$



Quartile Q_i

First quartile Q_1

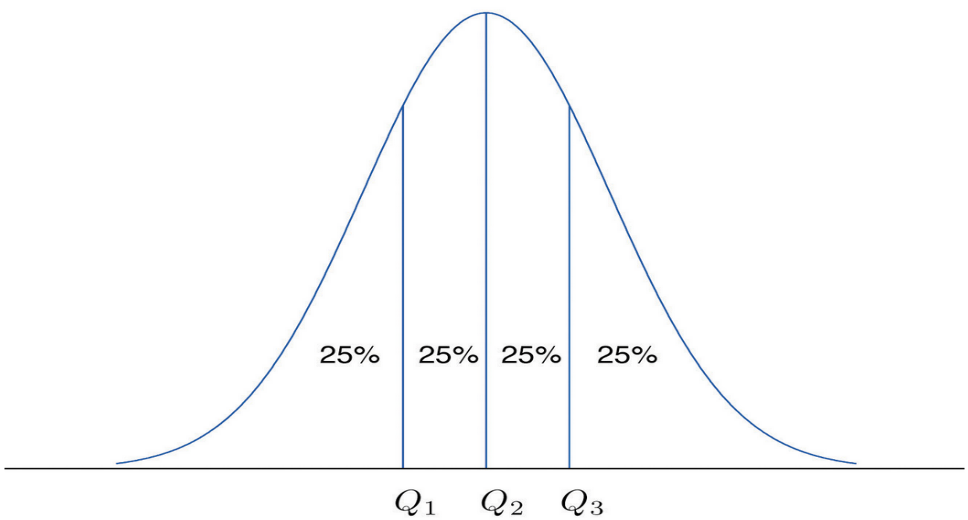
$$q_{0.25} = F^{-1}(0.25)$$

Median Q_2

$$q_{0.5} = F^{-1}(0.5)$$

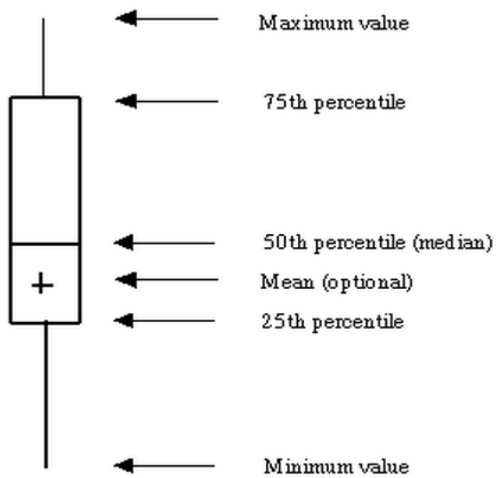
Third quartile Q_3

$$q_{0.75} = F^{-1}(0.75)$$

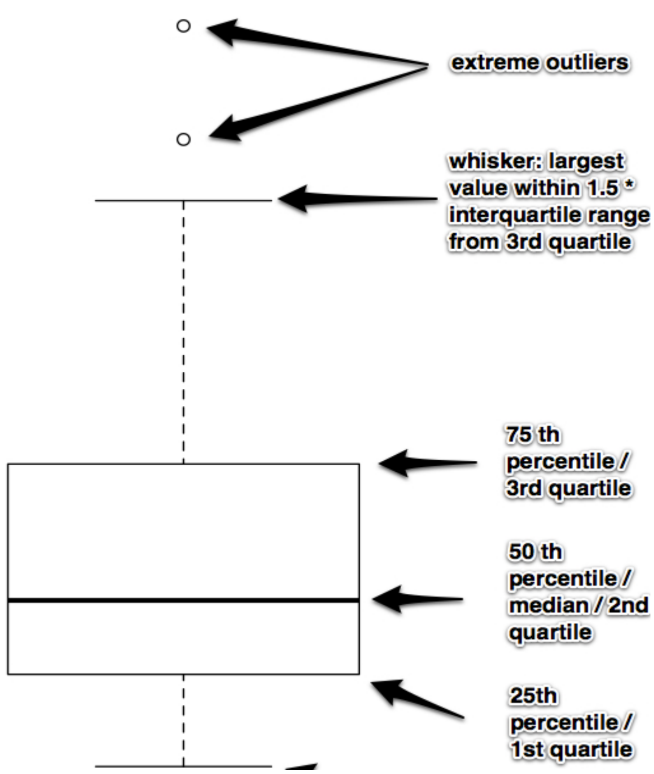


Box plot

Box plot



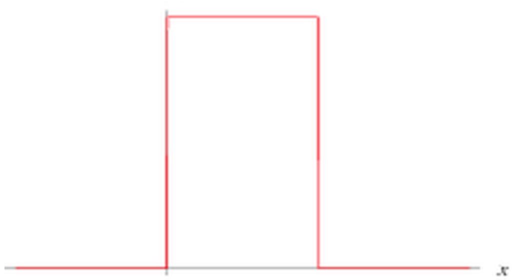
Modified box plot

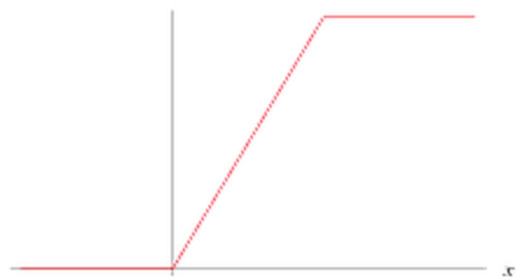


Example - CDF of $U(0,1)$

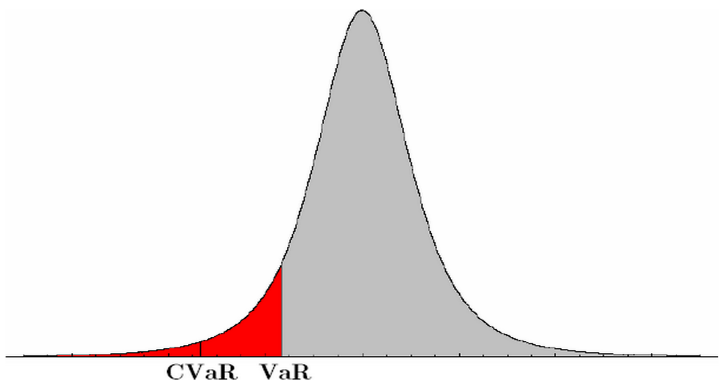
$$f(x) = \begin{cases} 0 & \text{for } x \le 0 \\ 1 & \text{for } 0 \le x \le 1 \\ 0 & \text{for } x \ge 1 \end{cases} \quad \overset{\text{Integrate}}{\Rightarrow} \quad F(x) = \begin{cases} 0 & \text{for } x \le 0 \\ x & \text{for } 0 \le x \le 1 \\ 1 & \text{for } x \ge 1 \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{for } x \le 0 \\ x & \text{for } 0 \le x \le 1 \\ 1 & \text{for } x \ge 1 \end{cases} \quad \overset{\text{Differentiate}}{\Rightarrow} \quad f(x) = \begin{cases} 0 & \text{for } x \le 0 \\ 1 & \text{for } 0 \le x \le 1 \\ 0 & \text{for } x \ge 1 \end{cases}$$





VaR and CVaR



$\boxed{5\%\ VaR}$

Let X be the P&L of a portfolio for a certain investment period. We can think 100 different possible outcomes. Among these look at the 5-th worst outcome, that is the 5% VaR:

$$VaR_{5\%} = q^X_{0.05}$$

$\boxed{5\%\ CVaR}$

Let X be the P&L of a portfolio for a certain investment period. We can think 100 different possible outcomes. Among these consider only the first five serious outcomes and compute the expected loss, that is the 5% $CVaR$:

$$CVaR_{5\%} = \mathbb{E}(X|X \leq VaR_{5\%})$$

Exponential distribution $Exp(\lambda)$

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } x \geq 0$$

Intuition First arrival time of Poisson point of intensity λ

Mean and variance $\frac{1}{\lambda}, \quad \frac{1}{\lambda^2}$

CDF $F(t) := P(T \leq t) = 1 - e^{-\lambda t} \quad \text{for } t \geq 0$

Survival function $\bar{F}(t) := 1 - F(t) = P(T > t) = e^{-\lambda t} \quad \text{for } t \geq 0$

Distribution	Random variable
$B(p)$	Flip a p-coin and check whether we have a head
$B(n,p)$	Flip a p-coin n times and count the number of heads
$Po(\lambda) \approx B(n,p)$	Flip a p-coin n times and count the number of heads, where $np = \lambda$ is fixed and $n \to \infty$
$Geo(p)$	Flip a p-coin until first head and count the number of flips
$Exp(\lambda) \approx \frac{1}{n} Geo(p)$	Flip a p-coin until first head and count the number of flips, where $np = \lambda$ is fixed and $n \to \infty$
$NB(r,p)$	Flip a p-coin until r-th head and count the number of flips

Distribution	Expectation	Variance
$B(p)$	p	pq
$B(n,p)$	np	npq
$Po(\lambda) \approx B(n,p)$	λ	λ
$Geo(p)$	$\frac{1}{p}$	$\frac{q}{p^2}$
$Exp(\lambda) \approx \frac{1}{n} Geo(p)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$NB(r,p)$	$\frac{r}{p}$	$\frac{rq}{p^2}$

$$\begin{aligned}
EX &= \int_0^\infty t\lambda e^{-\lambda t}dt = \int_0^\infty t(-e^{-\lambda t})'dt = \left[t(-e^{-\lambda t})\right]_0^\infty - \int_0^\infty (t)'(-e^{-\lambda t})dt \\
&= \int_0^\infty e^{-\lambda t}dt = \left[\frac{e^{-\lambda t}}{-\lambda}\right]_0^\infty = \frac{1}{\lambda}
\end{aligned}$$

$$\begin{aligned}
EX^2 &= \int_0^\infty t^2\lambda e^{-\lambda t}dt = \int_0^\infty t^2(-e^{-\lambda t})'dt = \left[t^2(-e^{-\lambda t})\right]_0^\infty - \int_0^\infty (t^2)'(-e^{-\lambda t})dt \\
&= 2\int_0^\infty te^{-\lambda t}dt = \frac{2}{\lambda}\int_0^\infty t\lambda e^{-\lambda t}dt = \frac{2}{\lambda}EX = \frac{2}{\lambda^2}
\end{aligned}$$

$$Var(X) = \frac{1}{\lambda^2}$$

Exponential approximation

If n is large, p is small, $np = \lambda$ is medium,

$$\frac{1}{n} Geo(p) \approx Exp(\lambda)$$

meaning, with $X \sim Geo(p)$, $Y \sim Exp(\lambda)$, for any $t \geq 0$

$$P\left(\frac{1}{n}X > t\right) \quad \approx \quad P(Y > t)$$

or

$$P\left(\frac{1}{n}X = t\right) \quad \approx \quad f_Y(t)dt$$

Exact mean match and approximate variance match

When n large, p small ($q \approx 1$), $np = \lambda$ medium	Exact		Approximate
Distribution of Number of p-coin flips to first head	$Geo(p)$	$-$	$-$
Distribution of Time to first head	$\frac{1}{n}Geo(p)$	$\approx$	$Exp(\lambda)$
(Exact) Mean match	$\frac{1}{np}$	$=$	$\frac{1}{\lambda}$
(Approximate) Variance match	$\frac{q}{(np)^2}$	$\approx$	$\frac{1}{\lambda^2}$

$$P\left(\frac{1}{n}X > t\right) = \sum_{i=nt+1}^{\infty} q^{i-1}p = q^{nt} = \left(1 - \frac{\lambda}{n}\right)^{nt} \quad \approx \quad e^{-\lambda t} = P(Y > t)$$

$$P\left(\frac{1}{n}X = t\right) = q^{nt-1}p = \lambda\left(1 - \frac{\lambda}{n}\right)^{nt-1} \underbrace{\frac{1}{n}}_{dt} \quad \approx \quad \lambda e^{-\lambda t}dt = f_Y(t)dt$$

Memoryless properties of geometric and exponential distribution

Geometric distribution

$$P(X > t|X > s) = \frac{q^t}{q^s} = q^{t-s} = P(X > t - s)$$

Exponential distribution

$$P(X > t|X > s) = \frac{e^{-\lambda t}}{e^{-\lambda s}} = e^{-\lambda(t-s)} = P(X > t - s)$$

Construction of Poisson point process $PPP(\lambda)$ using an $Exp(\lambda)$-coin

Number of ticks per year	n
p-coin flip at tick k	H with marker and T with no marker
Number of ticks to first head	$Geo(p)$
Time T_1 to first head	$\frac{1}{n} * Geo(p) \approx Exp(\lambda)$
Number of ticks to next head after i-th head	$Geo(p)$
Time T_{i+1} to next head after i-th head	$\frac{1}{n} * Geo(p) \approx Exp(\lambda)$

Construction of Poisson point process with intensity λ using an $Exp(\lambda)$-coin

[**Step 1**] Generate iid T_i from $Exp(\lambda)$.

[**Step 2**] Put a market at T_1, $T_1 + T_2$, $T_1 + T_2 + T_3, \cdots$.

Example - Joint PDF of independent random variables

The joint PDF of X and Y is given by

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & \text{for } 0 < x, y < \infty \\ 0 & \text{otherwise} \end{cases}$$

(1) Compute $P(X > 1, Y < 1)$.
(2) Compute $P(X < Y)$.
(3) $X \sim Exp(1)$ and $Y \sim Exp(2)$ are independent.

$$\begin{aligned}
P(X > 1, Y < 1) &= P(X > 1, 0 < Y < 1) \\
&= \int_1^\infty dx \int_0^1 dy \left[2e^{-x}e^{-2y}\right] \\
&= \int_0^1 dy \left[-2e^{-x}e^{-2y}\right]_{x=1}^{x=\infty} = \int_0^1 dy \left[2e^{-1}e^{-2y}\right] \\
&= \left[-e^{-1}e^{-2y}\right]_{y=0}^{y=1} = e^{-1}(1-e^{-2})
\end{aligned}$$

$$\begin{aligned}
P(X < Y) &= \int_0^\infty dx \int_x^\infty dy \left[2e^{-x}e^{-2y}\right] \\
&= \int_0^\infty dx \left[-e^{-x}e^{-2y}\right]_{y=x}^{y=\infty} = \int_0^\infty dx \left[e^{-3x}\right] \\
&= \left[-\frac{1}{3}e^{-3x}\right]_{x=0}^{x=\infty} = \frac{1}{3}
\end{aligned}$$

$$f(x,y) = 2e^{-x}e^{-2y}1(x \geq 0, y \geq 0) = \underbrace{e^{-x}1(x \geq 0)}_{Exp(1)} \cdot \underbrace{2e^{-2y}1(y \geq 0)}_{Exp(2)}$$

$$\Rightarrow \quad X \text{ and } Y \text{ are independent}$$

Example - Hike over Mt. Bukhan

During the hike over Mt. Bukhan I saw a warning sign saying that at a particular spot there are 2 mortality accidents per year on average. Calculate

(1) the probably that there is no mortality accident next one year.

(2) mean and variance of time that a mortality accident occur, starting from now.

Decompose 1 year into n ticks.

$1_{A_k} \sim B(p)$	Indicator of a mortality accident at tick k
$S_n = \sum_{k=1}^{n} 1_{A_k} \sim B(n,p)$	(Approx) Numb of mortality accident in a year
$S_n = \sum_{k=1}^{n} 1_{A_k} \sim B(n,p) \approx Po(\lambda)$	Numb of mortality accident in a year

With $S_n \sim B(n,p)$ and $X \sim Po(\lambda)$, $\lambda = np = 2$,

$$P(X = 0) = e^{-2} = 0.1353$$

$T \sim Exp(\lambda)$ Time that a mortality accident occur, starting from now

$$\mathbb{E}[T] = \frac{1}{\lambda} \quad \text{and} \quad Var(T) = \frac{1}{\lambda^2}$$

Thinning and merger

Thinning

[Step 1] Generate Poisson point process with intensity λ.

[Step 2] For each Poisson point we flip a α-coin independently.

[Step 3] If head, move this point to the up process.

[Step 4] If tail, move this point to the down process.

Then

(1) The up process is the Poisson point process with rate $\lambda\alpha$.
(2) The down process is the Poisson point process with rate $\lambda(1-\alpha)$.
(3) The up and down processes are independent.

Merger

[Step 1] Generate two indep Poisson point processes with intensity λ_1 and λ_2.

[Step 2] Merge these two.

Then

The merged process is the Poisson point process with rate $\lambda_1 + \lambda_2$.

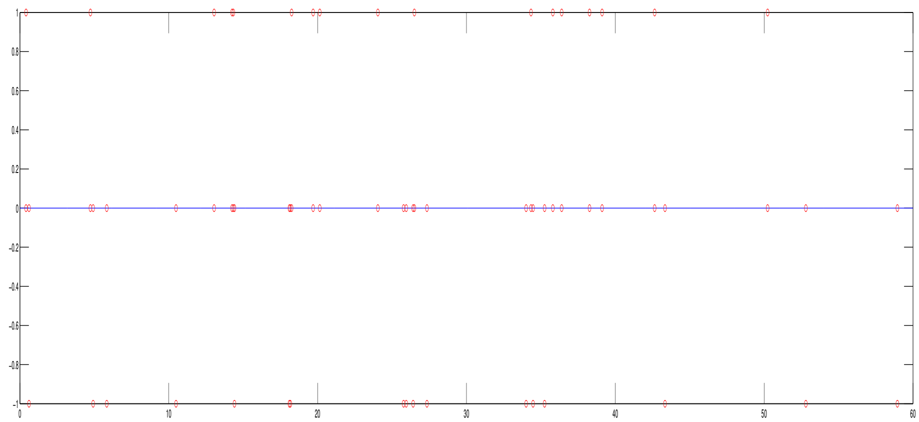


Figure 4: Poisson point process; thinning and merger.

```
clear all; close all; clc; rng('default')

la=0.5;

% Maximum range 0:n considered for Poisson point process
n=60;

% Thinning parameter
p=0.4;

% Poisson point process
x=random('exp',(1/la)*ones(n*la*4,1));
x=cumsum(x);
ind=find(x<=n,1,'last');
x=x(1:ind);
z=0:0.01:n;
plot(z,0,'-',x,zeros(size(x)),'or'); hold on;

% Thinning (Going up)
coin=rand(length(x),1);
up_ind=find(coin<=p);
down_ind=find(coin>p);
plot(x(up_ind),1,'or')

% Thinning (Going down)
plot(x(down_ind),-1,'or')
```

Theoretical backup for thinning and merger

Theoretical backup for thinning

Let X be the number of Poisson points in $[0,1]$ with intensity λ. For each Poisson point in $[0,1]$ we flip a α-coin independently to decide whether we move this point up or down. Let U be the number of the up points and let D be the number of the down points. Then

$$\begin{array}{ll} (1) & U \sim Po(\lambda\alpha) \\ (2) & D \sim Po(\lambda(1-\alpha)) \\ (3) & U \text{ and } D \text{ are independent} \end{array}$$

Theoretical backup for merger

$$Po(\lambda_1) \quad * \quad Po(\lambda_2) \quad = \quad Po(\lambda_1+\lambda_2)$$

$$\begin{aligned} P(U=u,D=d) &= P(U+D=u+d)P(U=u|U+D=u+d) \\ &= e^{-\lambda}\frac{\lambda^{u+d}}{(u+d)!}\binom{u+d}{u}\alpha^u(1-\alpha)^d \\ &= \underbrace{e^{-\lambda\alpha}\frac{(\lambda\alpha)^u}{u!}}_{U\sim Po(\lambda\alpha)} \quad \cdot \quad \underbrace{e^{-\lambda(1-\alpha)}\frac{(\lambda(1-\alpha))^d}{d!}}_{D\sim Po(\lambda(1-\alpha))} \end{aligned}$$

Theoretical backup for thinning is a little counter-intuitive

Let X be $B(n,p)$ number of iid uniform points in $[0,1]$, instead of the number of Poisson points in $[0,1]$ with intensity λ. For each uniform point in $[0,1]$ we flip a α-coin independently to decide whether we move this point up or down. Let U be the number of the up points and let D be the number of the down points. Then

(1) $U \sim B(n,p\alpha)$

(2) $D \sim B(n,p(1-\alpha))$

(3) But, U and D are not independent

For $0 \le u \le n$,

$$\begin{aligned}
P(U=u) &= \sum_{k=u}^{n} P(U+D=k)P(U=u|U+D=k) \\
&= \sum_{k=u}^{n} \binom{n}{k} p^k q^{n-k} \binom{k}{u} \alpha^u (1-\alpha)^{k-u} \\
&= \sum_{k=u}^{n} \binom{n}{k} q^{n-k} \binom{k}{u} (p\alpha)^u (p(1-\alpha))^{k-u} \\
&= (p\alpha)^u \sum_{k=u}^{n} \binom{n}{k}\binom{k}{u} q^{n-k} (p(1-\alpha))^{k-u} \\
&= (p\alpha)^u \sum_{k=u}^{n} \underbrace{\frac{\binom{n}{k}\binom{k}{u}}{\binom{n-u}{k-u}}}_{\binom{n}{u}} \binom{n-u}{k-u} q^{n-k} (p(1-\alpha))^{k-u} \\
&= \binom{n}{u} (p\alpha)^u \sum_{k=u}^{n} \binom{n-u}{k-u} q^{n-k} (p(1-\alpha))^{k-u} \\
&= \binom{n}{u} (p\alpha)^u (q+p(1-\alpha))^{n-u} = \binom{n}{u} (p\alpha)^u (1-p\alpha)^{n-u}
\end{aligned}$$

$$\Rightarrow \quad U \quad \sim \quad B(n,p\alpha) \quad \text{and by the same token} \quad D \quad \sim \quad B(n,p(1-\alpha))$$

If U and D are independent $\Rightarrow$ $0 \le U+D \le 2n$ since $0 \le U \le n$ and $0 \le D \le n$

$\Rightarrow$ Contradiction

Example - Min over independent exponential random variables

Let X_i be $Exp(\lambda_i)$. Suppose they are independent.

(1) What is the distribution of $\min\{X_1, X_2\}$?
(2) More generally, what is the distribution of $\min\{X_i, 1 \leq i \leq n\}$?
(3) What are the mean and variance of $\min\{X_i, 1 \leq i \leq n\}$?

Generate n independent Poisson point processes with intensity λ_i.

X_i	First arrival time of i^{th} Poisson point process
$\min\{X_1, X_2\}$	First arrival time of merged process merging first two
$\min\{X_i, 1 \leq i \leq n\}$	First arrival time of merged process merging all n

$$\Rightarrow \quad \min\{X_i, 1 \leq i \leq n\} \sim Exp\left(\sum_{i=1}^{n} \lambda_i\right)$$

$$\text{Mean} \quad \frac{1}{\sum_{i=1}^{n} \lambda_i}, \qquad \text{Variance} \quad \frac{1}{(\sum_{i=1}^{n} \lambda_i)^2}$$

Example - Waiting time to play tennis

An athletic facility has 5 tennis courts. Pairs of players arrive at the courts and use a court for an exponentially distributed time with mean 40 minutes, i.e., the play time is iid $Exp(\lambda)$, $\lambda^{-1} = 40$ (in minutes). When me and my partner arrive, we find all courts busy and 2 other pairs waiting in queue. What is the expected waiting time to get a court?

$X_1 \sim Exp(5\lambda)$	Time that first couple finish their game and leave the tennis court
	First couple in queue start to play
$X_2 \sim Exp(5\lambda)$	Time that second couple finish their game and leave the court measured from time that first couple leave
	Second couple in queue start to play
$X_3 \sim Exp(5\lambda)$	Time that third couple finish their game and leave the court measured from time that second couple leave
	Now, we just get a court!

$T = X_1 + X_2 + X_3$ Waiting time that we get a court, where X_i are iid $Exp(5\lambda)$

$$\mathbb{E}[T] = \sum_{i=1}^{3} \mathbb{E}[X_i] = \frac{3}{5\lambda}$$

Example - Queue at the bank - Part 1

When I enter the bank, there are already two people in line waiting for the service and I join the queue next to Soyoung, the last person in the line. There are four service desks and we assume the service time is iid $Exp(\lambda)$, $\lambda^{-1} = 10$ (in minutes). Calculate

the mean and variance of time T that I get serviced, starting from now.

$X_1 \sim Exp(4\lambda)$	Time that first person leaves the service desk
	First person in queue start to get one's service
$X_2 \sim Exp(4\lambda)$	Time that second person leaves the service desk
	measured from time that first person leaves the service desk
	Soyoung start to get her service
$X_3 \sim Exp(4\lambda)$	Time that third person leaves the service desk
	measured from time that second person leaves the service desk
	I start to get my service
$X_4 \sim Exp(\lambda)$	Time that my service is completed
	measured from time that third person leaves the service desk
	Now, I am leaving!

$T = X_1 + X_2 + X_3 + X_4$ Time that I get serviced, where X_i are independent

$$\mathbb{E}[T] = \sum_{i=1}^{4} \mathbb{E}[X_i] = \frac{3}{4\lambda} + \frac{1}{\lambda} \quad \text{and} \quad Var(T) = \sum_{i=1}^{4} Var(X_i) = \frac{3}{(4\lambda)^2} + \frac{1}{\lambda^2}$$

Example - Queue at the bank - Part 2

When I enter the bank, there are already two people in line waiting for the service and I join the queue next to Soyoung, the last person in the line. There are four service desks and we assume the service time is iid $Exp(\lambda)$, $\lambda^{-1} = 10$ (in minutes). Calculate

the probability that I leave the bank before Soyoung.

When Soyoung start to get her service, due to the memoryless property of the exponential distribution each of four in service has the equal chance of leaving first and hence she cannot leave first with probability 3/4. Then, I start to get my service. Again, by the memoryless property she and I have equal chance of leaving first among two. Therefore, the probability that I leave the bank before she leaves, is

$$\frac{3}{4} \times \frac{1}{2} = \frac{3}{8}$$