

变换 (针对连续型)

Lecture 7: Transformations

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

April 20, 2023

Outline

- 1 Change of Variables
- 2 Convolutions
- 3 Order Statistics
- 4 Beta-Binomial Conjugacy
- 5 Gamma Distribution
- 6 Dirichlet-Multinomial Conjugacy
- 7 Application Case: Bayesian Ranking

Outline

- 1 Change of Variables
- 2 Convolutions
- 3 Order Statistics
- 4 Beta-Binomial Conjugacy
- 5 Gamma Distribution
- 6 Dirichlet-Multinomial Conjugacy
- 7 Application Case: Bayesian Ranking

Change of Variables in One Dimension

換元

Theorem

Let X be a continuous r.v. with PDF f_X , and let $Y = g(X)$, where g is differentiable and strictly increasing (or strictly decreasing). Then the PDF of Y is given by

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$$

$Y = 2X$

$$f_Y(y) = f_X(x) \cdot \frac{1}{2}$$

where $x = g^{-1}(y)$. The support of Y is all $g(x)$ with x in the support of X .

Proof

① W.L.O.G... Let g be a strictly increasing function.

② We consider the CDF of Y .

$\forall y \in \mathbb{R}$.

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(\underline{g(x)} \leq y) \quad (y = g(x)) \\ &= P(X \leq g^{-1}(y)) \quad \stackrel{\Rightarrow X = g^{-1}(y)}{\quad} \\ &= F_X(g^{-1}(y)). \quad \stackrel{x \in \mathbb{R}}{=} \underline{F_X(x)}. \end{aligned}$$

\Rightarrow Chain rule, PDF of Y .

$$f_Y(y) = F_Y'(y) = \frac{\partial F_X(x)}{\partial x} \cdot \frac{dx}{dy} \quad \left. \right\} = f_X(x) \Big| \frac{dx}{dy}$$

③ If $g \downarrow$, $f_Y(y)$

$$= f_X(x) \left(-\frac{dx}{dy} \right) = f_X(x) \cdot \frac{dx}{dy}$$

Example: Log-Normal PDF

① $X = \log Y$, $X \sim N(0, 1)$, PDF of Y .

$$Y = g(x) = e^x \quad [y = e^x > 0, \Rightarrow x = \underline{\log y} \\ \Rightarrow \frac{dx}{dy} = \frac{1}{y}]$$

② $f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$
 $= f_X(x) \cdot \frac{1}{y} = f_X(\underline{\log y}) \cdot \frac{1}{y}$

③ $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$,
 $\Rightarrow f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\log y)^2} \cdot \frac{1}{y}, y > 0$

Change of Variables

Theorem

g : 可逆变换

Let $\mathbf{X} = (X_1, \dots, X_n)$ be a continuous random vector with joint PDF $f_{\mathbf{X}}(x)$, and let $\mathbf{Y} = g(\mathbf{X})$ where g is an invertible function from \mathbb{R}^n to \mathbb{R}^n . Let $y = g(\mathbf{x})$ and suppose that all the partial derivatives $\frac{\partial x_i}{\partial y_j}$ exists and are continuous, so we can form the **Jacobian matrix**

$$\frac{\partial \mathbf{x}}{\partial \mathbf{y}} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix}$$

$$\left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| = \boxed{\left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right|}$$

Also assume that the determinant of the Jacobian matrix is never 0. Then the joint PDF of \mathbf{Y} is

$$f_{\mathbf{Y}}(y) = f_{\mathbf{X}}(x) \left| \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right| \rightarrow \boxed{\left| \det \left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}} \right) \right|}$$

Jacobian or not

① Discrete r.v.s. $X, Y \geq 0, Y = X^3$

No Jacobian.

$$\underline{P(Y=y)} = P(X=y^{\frac{1}{3}})$$

② Continuous r.v.s.

$X, Y \geq 0, Y = X^3$

$$X = Y^3 \quad (\quad x = y^3 \quad \frac{dx}{dy} = 3y^2 \quad)$$

$$Y \geq 0, f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$$

$$= f_X(y^{\frac{1}{3}}) \cdot \frac{1}{3}y^{-\frac{2}{3}}$$

Box-Muller

$$\textcircled{1} \quad (X, Y) = g(U, T):$$

反三角轉換

$$\frac{\partial g(u,t)}{\partial (X,Y)}$$

$$X = \sqrt{t} \cos u$$

$$Y = \sqrt{t} \sin u$$

$$X^2 + Y^2 = 2t$$

$$t = \frac{1}{2}(X^2 + Y^2)$$

$$u = \arctan \frac{Y}{X}$$

$$f_{X,Y}(x,y) = f_{U,T}(u,t) \cdot \frac{1}{\left| \frac{\partial (X,Y)}{\partial (u,t)} \right|}$$

$$\frac{\partial (X,Y)}{\partial (u,t)}$$

$$\textcircled{2} \quad f_{U,T}(u,t) = f_U(u) \cdot f_T(t) = \frac{1}{2\pi} \cdot e^{-t} \quad u \in (0, 2\pi), t > 0$$

Let $U \sim \text{Unif}(0, 2\pi)$, and let $T \sim \text{Expo}(1)$ be independent of U . Define $X = \sqrt{2T} \cos U$ and $Y = \sqrt{2T} \sin U$. Find the joint PDF of (X, Y) . Are they independent? What are their marginal distributions?
 $\Rightarrow X, Y \stackrel{i.i.d.}{\sim} N(0, 1)$

$\textcircled{3}$. Jacobian.

$$\frac{\partial (X,Y)}{\partial (u,t)} = \begin{pmatrix} \frac{\partial X}{\partial u} & \frac{\partial X}{\partial t} \\ \frac{\partial Y}{\partial u} & \frac{\partial Y}{\partial t} \end{pmatrix} = \begin{pmatrix} -\sqrt{2t} \sin u & \frac{1}{\sqrt{2t}} \cos u \\ \sqrt{2t} \cos u & \frac{1}{\sqrt{2t}} \sin u \end{pmatrix}$$

$$\det \left(\frac{\partial (X,Y)}{\partial (u,t)} \right) = -\sin^2 u - \cos^2 u = -1 \quad (\therefore | = 1)$$

$$\textcircled{4} \quad f_{X,Y}(x,y) = \frac{1}{2\pi} e^{-t}, \quad | = \frac{1}{2\pi} e^{-\frac{1}{2}(X^2+Y^2)},$$

Solution

⑤ $f_{X,Y}(x,y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$

$x, y \in \mathbb{R}$

$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$

$\underline{h(x)} \cdot \underline{g(y)}$

$\Rightarrow X$ and Y are independent.

无相关

$$X \sim N(0,1)$$

$$Y \sim N(0,1)$$

Bivariate Normal Joint PDF

$$1^{\circ} \quad X, Y \sim \text{i.i.d. } N(0, 1)$$

一维分布线性组合 \Rightarrow 多元
正态 \star

$$\begin{cases} Z = X \\ W = \rho X + \sqrt{1-\rho^2} Y \end{cases}$$

$f_{Z,W}(z,w)$
~ Bivariate
Normal.
 $gm(z,w) = \rho$

$$2^{\circ} \quad (Z, W) = g(X, Y)$$

$$f_{Z,W}(z,w) = f_{X,Y}(x,y) \left| \frac{\partial(x,y)}{\partial(z,w)} \right|$$

$$3^{\circ} \quad \text{Jacobian} \quad \begin{cases} Z = X \\ W = \rho X + \sqrt{1-\rho^2} Y \end{cases} \Rightarrow \begin{cases} X = Z \\ Y = \frac{1}{\sqrt{1-\rho^2}}W - \frac{\rho}{\sqrt{1-\rho^2}}Z \end{cases}$$

$$\frac{\partial(x,y)}{\partial(z,w)} = \begin{pmatrix} \frac{\partial x}{\partial z} & \frac{\partial x}{\partial w} \\ \frac{\partial y}{\partial z} & \frac{\partial y}{\partial w} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{pmatrix}$$

$$\det(\downarrow) = \frac{1}{\sqrt{1-\rho^2}}$$

Bivariate Normal Joint PDF

$$\begin{cases} x = z \\ y = \frac{1}{\sqrt{1-\rho^2}}w - \frac{\rho}{\sqrt{1-\rho^2}}z \end{cases}$$

$$4^o: f_{Z,W}(z,w) = f_{X,Y}(x,y) \cdot \frac{1}{\sqrt{1-\rho^2}}$$

$$= f_X(x) \cdot f_Y(y) \cdot \frac{1}{\sqrt{1-\rho^2}}$$

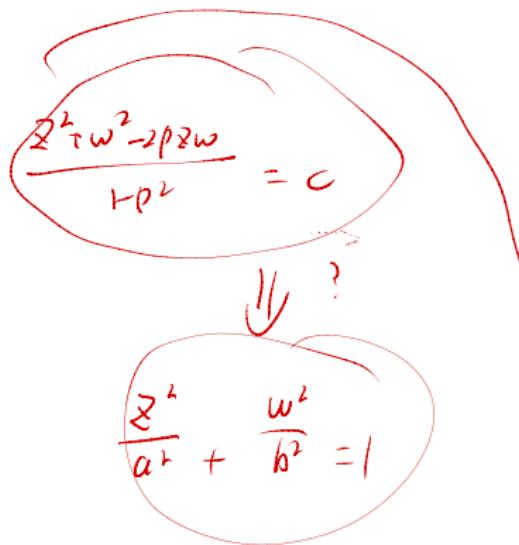
$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \cdot \frac{1}{\sqrt{1-\rho^2}}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}(x^2+y^2)}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}[z^2 + (\frac{1}{\sqrt{1-\rho^2}}w - \frac{\rho}{\sqrt{1-\rho^2}}z)^2]}$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(z^2 + w^2 - 2\rho zw)}$$

$z, w \in \mathbb{R}$.



Outline

- 1 Change of Variables
- 2 Convolutions
- 3 Order Statistics
- 4 Beta-Binomial Conjugacy
- 5 Gamma Distribution
- 6 Dirichlet-Multinomial Conjugacy
- 7 Application Case: Bayesian Ranking

卷积

Convolution Sums and Integrals

Theorem

MGF : 卷积 \Rightarrow 相乘

If X and Y are independent discrete r.v.s, then the PMF of their sum

$$\underbrace{T = X + Y \text{ is}}_{P(T=t)} \quad P(T=t) \stackrel{\text{defn}}{=} \sum_x P(X+Y=t | X=x) \cdot P(X=x)$$

$$P(T=t) = \sum_x \underbrace{P(Y=t-x)}_{\substack{= \sum_y P(Y=t-y | X=x) \\ \cdot P(X=x)}} P(X=x)$$

$$= \sum_y P(X=t-y) P(Y=y) = \sum_x \underbrace{P(Y=t-x)}_{\substack{= \sum_y P(Y=t-y | X=x) \\ \cdot P(X=x)}} P(X=x)$$

If X and Y are independent continuous r.v.s, then the PDF of their sum $T = X + Y$ is

$$f_T(t) = \int_{-\infty}^{\infty} f_Y(t-x) f_X(x) dx = f_Y(t) * f_X(t)$$

$$= \int_{-\infty}^{\infty} f_X(t-y) f_Y(y) dy = f_X(t) * f_Y(t)$$

Proof ①. For continuous r.v.s X, Y .

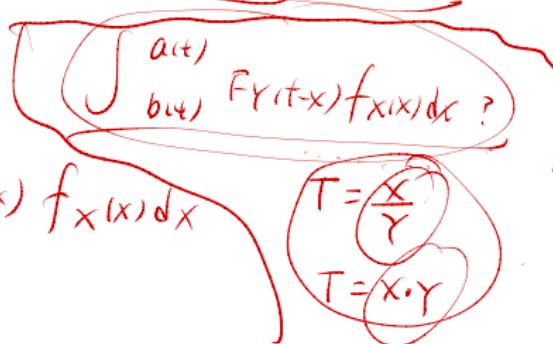
$$\begin{aligned} F_T(t) &= P(X+Y \leq t) \stackrel{\text{LorP}}{=} \int_{-\infty}^{\infty} P(X+Y \leq t | X=x) f_{X(x)} dx \\ &= \int_{-\infty}^{\infty} \underbrace{P(Y \leq t-x | X=x)}_{X, Y \text{ independent}} f_{X(x)} dx \end{aligned}$$

$$= \int_{-\infty}^{\infty} \underbrace{P(Y \leq t-x)} f_{X(x)} dx$$

$$= \int_{-\infty}^{\infty} \underbrace{F_Y(t-x)} f_{X(x)} dx$$

Diff wrt. t

$$f_T(t) = \int_{-\infty}^{\infty} f_Y(t-x) f_{X(x)} dx$$



Exponential Convolution

$$\begin{aligned} t &= x+y \\ v &= x \end{aligned} \Rightarrow \begin{cases} x = v \\ y = t-v \end{cases}$$

$$J = \begin{vmatrix} \frac{\partial x}{\partial t} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial t} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1$$

$$T = X+Y, V = X$$

$$(T, V) = g(X, Y)$$

$$\Rightarrow f_{T,V}(t,v) = f_{X,Y}(x,y) \cdot |J| = f_X(x) \cdot f_Y(y) \cdot 1 = f_X(v) f_Y(t-v).$$

Let $\underbrace{X, Y}_{\text{i.i.d.}} \sim \text{Expo}(\lambda)$. Find the distribution of $T = X + Y$.

$$\Rightarrow f_T(t) = \int_{-\infty}^{\infty} f_{T,V}(t,v) dv = \int_{-\infty}^{\infty} f_X(v) f_Y(t-v) dv$$

$$\begin{aligned} \text{if } t \geq 0 \quad f_T(t) &= \int_{-\infty}^{\infty} f_X(x) f_Y(t-x) dx \\ &= \int_0^t \lambda e^{-\lambda(t-x)} \cdot \lambda e^{-\lambda x} dx = \lambda^2 t e^{-\lambda t}. \end{aligned}$$

Outline

- 1 Change of Variables
- 2 Convolutions
- 3 Order Statistics
- 4 Beta-Binomial Conjugacy
- 5 Gamma Distribution
- 6 Dirichlet-Multinomial Conjugacy
- 7 Application Case: Bayesian Ranking

Order Statistics

有相关性
(互不独立关系)

$X_{(1)}, \dots, X_{(n)}$ dependent

$X_{(1)} = 100; X_{(n)} \geq 100$

$\circled{X_{(j)} \geq 100, j=2, \dots, n}$

Definition

For r.v.s X_1, X_2, \dots, X_n , the *order statistics* are the random variables $X_{(1)}, \dots, X_{(n)}$, where

$$X_{(j)} = \underline{f(X_1, \dots, X_n)}$$

$$\underline{X_{(1)} = \min(X_1, \dots, X_n)},$$

$\underline{X_{(2)}}$ is the second-smallest of X_1, \dots, X_n ,

\vdots

$\underline{X_{(n-1)}}$ is the second-largest of X_1, \dots, X_n ,

$$\underline{X_{(n)} = \max(X_1, \dots, X_n)}$$

Note that $\underline{X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}}$ by definition. We call $\underline{X_{(j)}}$ the jth order statistic.

第j顺序统计量

Order Statistics

X_1, \dots, X_n are i.i.d. Continuous r.v.s.

CDF F
PDF f

1^o. $X_{(n)} = \overbrace{\max(X_1, \dots, X_n)}$

$$\forall x \in \mathbb{R}, F_{X_{(n)}}(x) = P(X_{(n)} \leq x) = P(\max(X_1, \dots, X_n) \leq x)$$

$$= P(X_1 \leq x, \dots, X_n \leq x) = \underbrace{P(X_1 \leq x)} \cdots \underbrace{P(X_n \leq x)} = [F(x)]^n.$$

$$\Rightarrow f_{X_{(n)}}(x) = \underbrace{n(F(x))^{n-1}} \cdot f(x).$$

2^o. $X_{(1)} = \min(X_1, \dots, X_n)$

$$\forall x \in \mathbb{R}, F_{X_{(1)}}(x) = P(\underbrace{X_{(1)} \leq x}) = 1 - P(X_{(1)} > x)$$

$$= 1 - P(\min(X_1, \dots, X_n) > x) = 1 - (\underbrace{P(X_1 > x)} \cdots \underbrace{P(X_n > x)})^{1-F(x)}.$$

$$\Rightarrow f_{X_{(1)}}(x) = n(1-F(x))^{n-1} \cdot f(x).$$

CDF of Order Statistics

$$N \sim \text{Bin}(n, F(x))$$

$$P(N=k) = \frac{\binom{n}{k} \cdot F(x)^k \cdot (1-F(x))^{n-k}}{e^{nF(x)} \cdot (1-e^{-F(x)})^n}$$

$$\begin{aligned} F(x) &= P(X \leq x) \\ &= e^{-F(x)} \end{aligned}$$

Theorem

Let X_1, \dots, X_n be i.i.d. continuous r.v.s with CDF F . Then the CDF of the j th order statistic $X_{(j)}$ is

$$P(X_{(j)} \leq x) = \sum_{k=j}^n \binom{n}{k} F(x)^k (1 - F(x))^{n-k}.$$

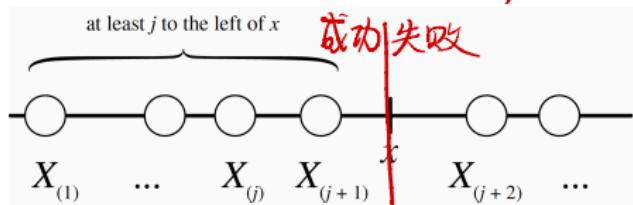
$$\sum_{k=j}^n \cdot P(N=k) = P(N \geq j)$$

Proof

① Given $x \in \mathbb{R}$, construct n independent Bernoulli trials.

For each trial i , the result is successful if r.v. X_i lands to the left of x ; such successful prob. $\underline{P(X_i < x)} = F(x)$

② Define a new rv N : # of X_i that land to the left of x .



$$N \sim \underline{\text{Bin}(n, F(x))}$$

③ $P(X_{(j)} \leq x) = P(\text{"at least } j \text{ of } X_1, \dots, X_n \text{ land to the left of } x")$

第 j 个 $\leq x$

\Rightarrow 至少 j 个 X 在 x 左边

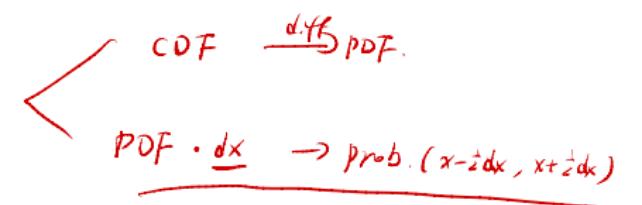
$$= P(N \geq j) = \sum_{k=j}^n \underline{P(N=k)}$$

$$= \sum_{k=j}^n \binom{n}{k} \cdot F(x)^k \cdot (1-F(x))^{n-k}.$$

Proof

PDF of Order Statistic

Two methods of the PDF



Theorem

Let X_1, \dots, X_n be i.i.d. continuous r.v.s with CDF F and PDF f . Then the marginal PDF of the j th order statistic $X_{(j)}$ is

$$f_{X_{(j)}}(x) = n \binom{n-1}{j-1} f(x) F(x)^{j-1} (1-F(x))^{n-j}.$$

第 j 个顺序统计量

Unif(0,1) $f_{X_{(j)}}(x) \cdot dx \sim \text{prob. } (X_{(j)} \in (x - \frac{1}{2}dx, x + \frac{1}{2}dx))$

$\sim \text{Beta}(j, n-j+1)$

Proof

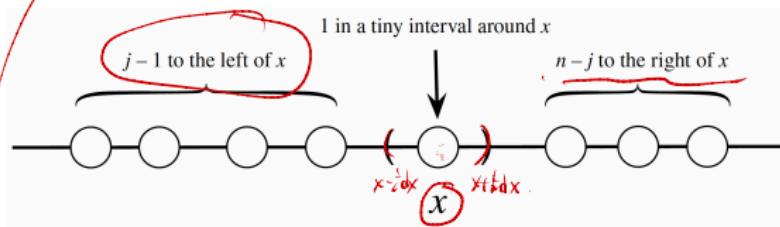
$$\textcircled{1} \quad \text{prob. } (\underline{X_{(j)} \in (x-\frac{1}{2}dx, x+\frac{1}{2}dx)}) \approx f_{X(j)}(x) dx$$

$$f_{X(j)}(x)$$

$$= nf(x) \cdot \binom{n-1}{j-1}$$

$$\cdot [F(x)]^{j-1}$$

$$\cdot [1-F(x)]^{n-j}$$



$$\underbrace{f(x) \cdot dx}_{\downarrow} \\ n \cdot f(x) dx$$

2^o. choose $j-1$ out of $n-1$ remaining r.v.s. , put on the left of x .

$$N^1 \sim \text{Bin}(n-1, F(x)) ;$$

$$p(N^1 = j-1) = \binom{n-1}{j-1} \cdot [F(x)]^{j-1} \cdot [1-F(x)]^{n-j}$$

$$\textcircled{3} \quad f_{X(j)}(x) \cdot dx \stackrel{?}{=} \lim_{dx \rightarrow 0} \frac{nf(x) \cdot dx \cdot \binom{n-1}{j-1} [F(x)]^{j-1} [1-F(x)]^{n-j}}{dx}$$

Example: Order Statistics of Uniforms

1^o $U_1, \dots, U_n \sim \text{i.i.d. Uniform}$. (U)

CDF $F_U(x) = x \quad (0 \leq x \leq 1)$

PDF $f_U(x) = 1 \quad (0 \leq x \leq 1)$

2^o $U_{(1)}, \dots, U_{(n)}$ order statistic.



$$f_{U_{(j)}}(x) = n \binom{n-1}{j-1} f_U(x) \underbrace{[F_U(x)]^{j-1}}_{\sim} [1 - F_U(x)]^{n-j}$$
$$= n \binom{n-1}{j-1} \cdot 1 \cdot x^{j-1} (1-x)^{n-j}$$

$U_{(j)} \sim \text{Beta}(j, n-j)$

β if $\Rightarrow = \frac{n!}{(n-j)! (j-1)!} \cdot x^{j-1} (1-x)^{n-j}$

Example: Order Statistics of Uniforms

Related Identity

Application : RL (Thompson Sampling)

$$\textcircled{1} \quad U_1, \dots, U_n \sim \text{i.i.d. unif}(0,1).$$

强化学习

$$\xrightarrow{p}$$

分析 Thompson采样性能

n independent Bernoulli trials. " $U_i < p$ " success
 $\Pr(U_i < p) = p$

Theorem

For $0 < p < 1$, and nonnegative integer k , we have

$$\sum_{j=0}^k \binom{n}{j} p^j (1-p)^{n-j} = \frac{n!}{k!(n-k-1)!} \int_p^1 x^k (1-x)^{n-k-1} dx$$

\textcircled{2} N : # of successful trials. $N \sim \text{Bin}(n, p)$.

$$\begin{aligned} \textcircled{3} \quad \text{LHS} : \underline{\Pr(N \leq k)} &= \Pr(\text{"at most } k \text{ of } U_1, \dots, U_n \text{ are left of } p") \\ &= \sum_{j=0}^k \underline{\Pr(N=j)} \\ &= \underline{\Pr(U_{(k+1)} > p)} = \underline{\int_p^1 f_{U_{(k+1)}}(x) dx} \end{aligned}$$

Proof

Outline

- 1 Change of Variables
- 2 Convolutions
- 3 Order Statistics
- 4 Beta-Binomial Conjugacy
- 5 Gamma Distribution
- 6 Dirichlet-Multinomial Conjugacy
- 7 Application Case: Bayesian Ranking

Beta Distribution $\stackrel{!}{\circ}$ $a=b=1$, $f(x) = \frac{1}{\beta(a,b)} = \frac{1}{\beta(1,1)}$ is a constant.

β 分布

$$\int_0^1 f(x) dx = 1 \Rightarrow \frac{1}{\beta(1,1)} = 1 \Rightarrow \int_0^1 f(x) dx = 1$$

motivation 在 $(0,1)$ 区间上推广 unif

Definition

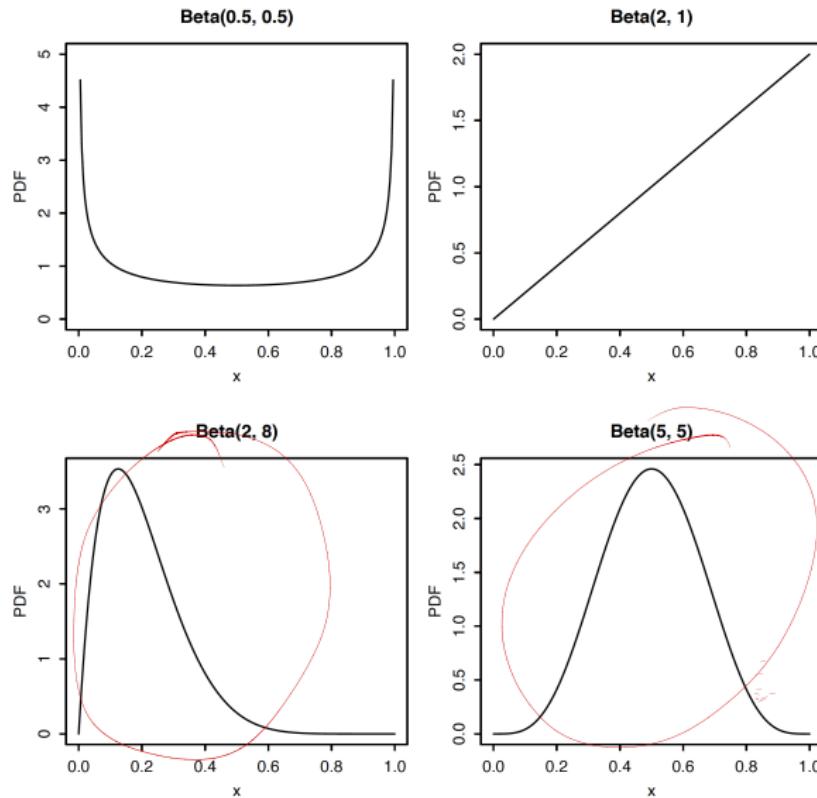
An r.v. X is said to have the *Beta distribution* with parameters a and b , $a > 0$ and $b > 0$, if its PDF is

$$f(x) \propto x^{a-1} (1-x)^{b-1}$$

$$f(x) = \frac{1}{\beta(a,b)} x^{a-1} (1-x)^{b-1}, \quad 0 < x < 1,$$

where the constant $\beta(a,b)$ is chosen to make the PDF integrate to 1. We write this as $X \sim \text{Beta}(a,b)$. Beta distribution is a generalization of uniform distribution.

PDF of Beta Distribution



Beta Integral

β 积分

β 函数

$$\beta(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx.$$

$$a, b \in \mathbb{R} \quad \beta(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad \Gamma(a) = (a-1)!$$

$$a, b \in \mathbb{Z} \quad \beta(a, b) = \frac{(a-1)!(b-1)!}{(a+b-1)!}$$

Story: Bayes' billiards

$$\textcircled{2} \quad P(Y=k \mid X=x)$$

$$\textcircled{3} \quad \int_0^1 P(Y=k \mid X=x) \cdot f_X(x) \cdot dx = \frac{\binom{n}{k} x^k (1-x)^{n-k}}{P(Y=k)}$$

Show without using calculus that for any integers k and n with $0 \leq k \leq n$,

$$\int_0^1 \binom{n}{k} x^k (1-x)^{n-k} dx = \frac{1}{n+1}.$$

x ~ $\text{Bin}(n, x)$ $P(Y=k)$

复合随机分布
参数为变量

$$\textcircled{1} \quad Y \sim \text{Bin}(n, x), \quad x \text{ is a r.v. } \sim \text{unif}(0, 1)$$

$f_X(x) = 1$

$Y \mid X=x \sim \text{bin}(n, x)$

Solution

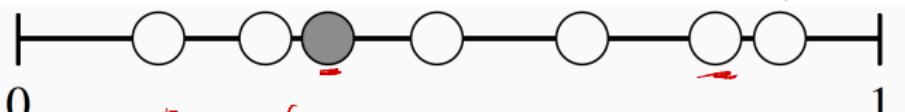
Story 1: $n+1$ balls; n white balls, 1 gray ball.

Randomly put each ball $\rightarrow [0, 1]$, s.t. position of balls

$\sim \text{i.i.d. uniform}$

Y : # of white balls to the left of the gray ball.

Y : discrete r.v. $\in \{0, 1, \dots, n\}$.



PMF of Y : $P(Y=k) ? k \in \{0, 1, \dots, n\}$

① Conditioning on the position of gray ball B , $\sim \text{uniform}$

$$f_B(p) = 1 = 1, 0 < p < 1$$

② Conditioning on $B=p$, $Y|B=p \sim \text{Bin}(n, p)$

$$\text{④ } P(Y=k) \stackrel{\text{LDT}}{=} \int_0^1 P(Y=k|B=p) f_B(p) dp = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dp$$

Solution

Story 2: ① $n+1$ white balls $\rightarrow [0,1]$

position $\sim \text{unif}(0,1)$ iid.

② choose one ^{white} ball at random (equal prob.)
and paint it gray

③ Y : # of white balls to the left of
the gray ball.



$$\underline{P(Y=k)} = ?$$
$$\frac{1}{n+1} \checkmark$$

Solution

$$\int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dp = \frac{1}{n+1}.$$

$$k = a - 1$$

$$n-k=b-l$$

$$a, b > 0$$

integer

$$a = k+1 \quad ; \quad b = n-k+1 \quad ; \quad n = a+b-2 \quad ;$$

$$P = X$$

$$\Rightarrow \int_0^1 \left(\frac{a+b-2}{a-1} \right) x^{a-1} (1-x)^{b-1} dx = \frac{1}{a+b-1}$$

$$\Rightarrow \int_0^1 x^{a-1} (1-x)^{b-1} dx = \frac{1}{(a+b-2) \cdot (a+b-1)}$$

$$= \frac{(a-1)! (b-1)!}{(ab-1)!} = \binom{a+b-1}{a, b}$$

Story: Beta-Binomial Conjugacy 共轭

$$\hat{P} = \frac{k}{n}; \quad \begin{array}{l} n \text{ tosses.} \\ k \text{ lands Heads.} \end{array}$$

$$n = 10^{10}; \quad k = \frac{1}{2} \times 10^{10}; \quad \hat{P} = \frac{1}{2}.$$

We have a coin that lands Heads with probability p , but we don't know what p is. Our goal is to infer the value of p after observing the outcomes of n tosses of the coin. The larger that n is, the more accurately we should be able to estimate p .

$$\underline{n=5} \quad \underline{k=5}; \quad \hat{P}=1 \quad ?$$

Bayesian Inference

X

贝叶斯统计推断方法
(Bayesian Statistics)

- Treats all unknown quantities as random variables.
- In the Bayesian approach, we would treat the unknown probability p as a random variable and give p a distribution.
- This is called a **prior distribution**, and it reflects our uncertainty about the true value of p before observing the coin tosses.
- After the experiment is performed and the data are gathered, the prior distribution is updated using Bayes' rule; this yields the **posterior distribution**, which reflects our new beliefs about p .

$f(p)$

(与时俱进)

Story: Beta-Binomial Conjugacy

β 分布可以看成 $(0,1)$ 上
任意分布

① p : unknown, model it as a r.v. $\in [0, 1]$

Prior distribution: $p \sim \underline{\text{Beta}}(a, b)$.

② Data model: n tosses of coin.

X : # of heads.

X : 复合型分布
(条件二项分布)

$X | p=p \sim \text{Bin}(n, p)$ real number

r.v.

③

$f(p) = \underline{\text{prior distribution}}$ PDF of p .



$f(p | X=k) = \underline{\text{Posterior distribution}}$ PDF of p .

Story: Beta-Binomial Conjugacy

$P \sim \text{Beta}(\alpha, \beta)$

$$f(p|x=k) = \frac{\Pr(X=k|P)}{\Pr(X=k)} \cdot f(p) = \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\Pr(X=k)} \cdot \frac{1}{\beta(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1}$$

$$\Pr(X=k) \stackrel{\text{LoTP}}{=} \int_0^1 \Pr(X=k|p) f(p) dp = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} \cdot \frac{1}{\beta(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1} dp$$

分子与P无关(积分项积掉了)

- ④ $f(p|x=k)$ is a function of p. [every item that does not dependent on p can be regarded as a constant].

$$\begin{aligned} f(p|x=k) &= C \cdot \frac{p^k (1-p)^{n-k} \cdot p^{\alpha-1} (1-p)^{\beta-1}}{p^{k+\alpha-1} (1-p)^{n+k-\alpha-1}} \\ &\propto \frac{p^{k+\alpha-1} (1-p)^{n+k-\alpha-1}}{\beta(\alpha+k, n-k+\beta)} \end{aligned}$$

$$\Rightarrow P|X=k \sim \text{Beta}(\alpha+k, \beta+n-k)$$

Story: Beta-Binomial Conjugacy

$$P \sim \text{Beta}(\underline{a}, \underline{b})$$

Prior.

伪计数

Pseudo counts.

$$(\underline{a}, \underline{b})$$

of

prior successes

of prior failures

current
experience

data model.

Count k/k

$$\underline{X|P=p \sim \text{Bin}(n, p)}$$

X=k out of n
tosses.

$$P|X=k$$

$$\sim \text{Beta}(\underline{a+k}, \underline{b+n-k})$$

Posterior.

$$X=k$$

k successes

n-k failures

$$(\underline{a+k}, \underline{b+n-k})$$

$$\underline{X'=m} \quad (\underline{a+k+m}, \underline{b+n-k+n-m})$$

a
假想实验室中成功次数
b - · · · · · 失败 ·

Story: Beta-Binomial Conjugacy

- Furthermore, notice the very simple formula for updating the distribution of p .
- We just add the number of observed successes, k , to the first parameter of the Beta distribution.
- We also add the number of observed failures, $n - k$, to the second parameter of the Beta distribution.
- So a and b have a concrete interpretation in this context:
 - ▶ a as the number of prior successes in earlier experiments
 - ▶ b as the number of prior failures in earlier experiments
 - ▶ a, b : pseudo counts

Mean vs. Bayesian Average

$$\left[\begin{array}{l} Y \sim \text{Beta}(a, b) \\ E(Y) = \frac{a}{a+b} \end{array} \right]$$

- Infer the value of p (probability of coin lands heads)
- Observed k heads out of n tosses of the coin
- Mean: $\frac{k}{n}$ P的后验概率期望
- Bayesian Average: $E(p|X=k) = \frac{a+k}{a+b+n}$
- Suppose the prior distribution is $\text{Unif}(0,1)$: $a=1, b=1$
- Bayesian Average: $\frac{k+1}{n+2}$
- When $k=n$, we have: 1 (mean) vs. $\frac{n+1}{n+2}$ (Bayesian average)
 $n=2, 3, 4$
 $n \rightarrow \infty \rightarrow 1$

充分考虑样本数
据量大小对值
的影响

Story: Beta-Binomial Conjugacy

If we have a Beta prior distribution on p and data that are conditionally Binomial given p , then when going from prior to posterior, we don't leave the family of Beta distributions. We say that the Beta is the conjugate prior of the Binomial.

共轭先验

共轭对：

Beta - Binomial

Gamma - Poisson

Normal - Normal

Outline

- 1 Change of Variables
- 2 Convolutions
- 3 Order Statistics
- 4 Beta-Binomial Conjugacy
- 5 Gamma Distribution
- 6 Dirichlet-Multinomial Conjugacy
- 7 Application Case: Bayesian Ranking

Gamma Function

Beta: $(0, 1)$

Gamma: $(0, +\infty)$

Gamma Distribution (Motivation)

Prior distribution of unknown parameters $(0, \infty)$

Definition

The gamma function Γ is defined by

$$\text{tau} \stackrel{\rightarrow}{\circ} \Gamma(a) = \int_0^\infty x^a e^{-x} \frac{dx}{x},$$

for real numbers $a > 0$.

$$1 = \int_0^\infty \frac{1}{\Gamma(a)} x^{a-1} e^{-x} dx = \int_0^\infty \frac{f(x) dx}{\Gamma(a)}$$

↓ Gamma
(a, 1)
PDF.

Property of Gamma Function

- $\Gamma(a+1) = a\Gamma(a)$ for all $a > 0$
- $\Gamma(n) = (n-1)!$ if n is a positive integer.

$$\underline{\Gamma(n+1) = n\Gamma(n)}$$

正整数

$$\underline{\Gamma(1) = \int_0^\infty 1 \cdot e^{-x} dx = 1}$$

$$\Gamma(2) = 1$$

$$\Gamma(3) = 2, \dots \quad \Gamma(n) = (n-1)!$$

$$\begin{array}{c} 0!, 1!, 2! \\ \dots n! \\ (\frac{1}{2})! \end{array}$$

Gamma Distribution

$$a=1, f(y) = \frac{1}{\Gamma(1)} \cdot 1 \cdot e^{-y} \cdot \frac{1}{y}$$

指數分布的推广 $\Leftarrow = \frac{\lambda e^{-\lambda y}}{\text{Exp}(\lambda)}, y > 0.$

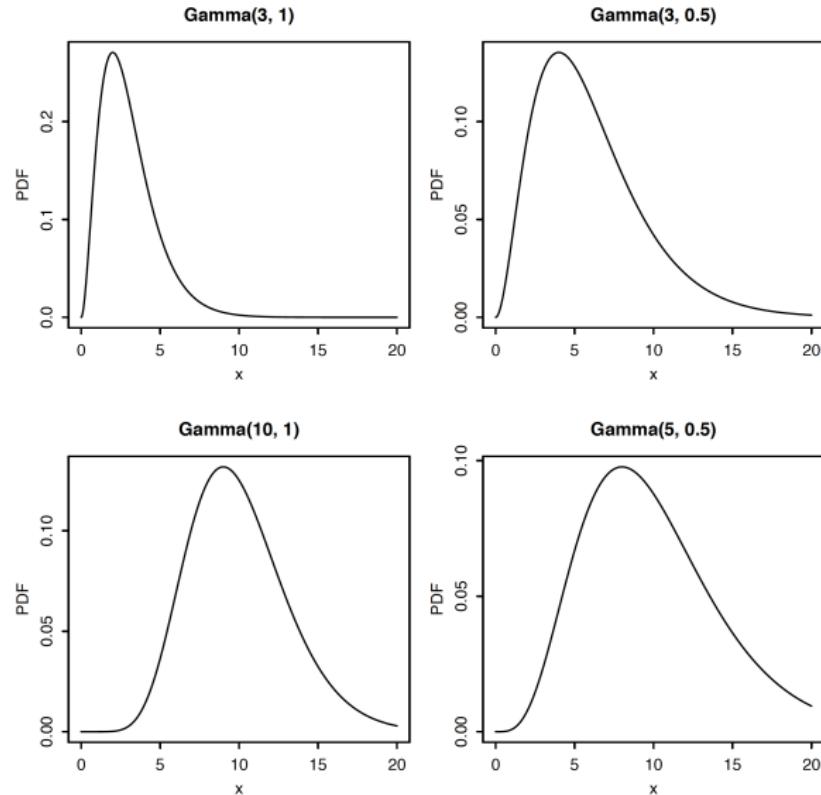
Definition

An r.v. Y is said to have the *Gamma distribution* with parameters a and λ , $a > 0$ and $\lambda > 0$, if its PDF is $(0, +\infty)$

$$f(y) = \frac{1}{\Gamma(a)} (\lambda y)^a e^{-\lambda y} \frac{1}{y}, y > 0.$$

We write $Y \sim \text{Gamma}(a, \lambda)$. Gamma distribution is a generalization of the exponential distribution.

PDF of Gamma Distribution



Moments of Gamma Distribution

$Y \sim \text{Gamma}(a, \lambda)$

$$E(Y) = \frac{a}{\lambda};$$

$$\text{Var}(Y) = \frac{a}{\lambda^2};$$

if $\lambda=0$,

$$E(Y) = \text{Var}(Y) = a$$

Pois

(Poisson - Gamma Duality)

if $X \sim \text{Pois}(\lambda)$, PMF $P(X=k) = \frac{\lambda^k e^{-\lambda}}{k!}$

if $X \sim \text{Gamma}(k+1, 1)$, PDF $f(x) = \frac{x^k e^{-x}}{k!}$

Gamma: Convolution of Exponential

Lec 5 Poisson process

Theorem

Let X_1, \dots, X_n be i.i.d. $\text{Expo}(\lambda)$. Then

$$X_1 + \dots + X_n \sim \text{Gamma}(n, \lambda).$$

$$\underline{M_{X(t)}}$$

MGF 相同

$$\underline{\text{MGF}} = \text{MGF}$$

分布相同

Proof

Beta-Gamma Connection

When we add independent Gamma r.v.s X and Y with the same rate λ , the total $X + Y$ has a Gamma distribution, the fraction $\frac{X}{X+Y}$ has a Beta distribution, and the total is independent of the fraction.

Story: Bank–post Office

While running errands, you need to go to the bank, then to the post office. Let $X \sim \text{Gamma}(a, \lambda)$ be your waiting time in line at the bank, and let $Y \sim \text{Gamma}(b, \lambda)$ be your waiting time in line at the post office (with the same λ for both). Assume X and Y are independent. What is the joint distribution of $T = X + Y$ (your total wait at the bank and post office) and $W = \frac{X}{X+Y}$ (the fraction of your waiting time spent at the bank)?

Story: Bank–post Office

$$T = X + Y ; \quad W = \frac{X}{X+Y}.$$

① $\begin{array}{l} t > 0 \\ w > 0 \end{array} \quad \begin{array}{l} t = xy \\ w = \frac{x}{xy} \end{array} \Rightarrow \begin{array}{l} x = tw \\ y = t(1-w) \end{array} \Rightarrow \frac{\partial(x,y)}{\partial(t,w)} = \begin{pmatrix} w & t \\ 1-w & -t \end{pmatrix}$

$$\det\left(\frac{\partial(x,y)}{\partial(t,w)}\right) = -t < 0. \quad \downarrow \quad | = t > 0.$$

② $f_{T,W}(t,w) = f_{X,Y}(x,y) \left| \frac{\partial(x,y)}{\partial(t,w)} \right| = f_X(x)f_Y(y).t. \quad (\underline{|} \cdot | = |\det(\cdot)|)$

$$\Rightarrow \frac{1}{\Gamma(a)} (\lambda x)^a e^{-\lambda x} \cdot \frac{1}{\Gamma(b)} (\lambda y)^b e^{-\lambda y} \cdot \frac{1}{\Gamma(c)} \cdot t \quad (\begin{array}{l} x = tw \\ y = t(1-w) \end{array})$$

$$= \frac{1}{\Gamma(a)} \cdot \frac{1}{\Gamma(b)} \cdot w^{a-1} (1-w)^{b-1} \cdot (\lambda t)^{a+b} e^{-\lambda t} \cdot \frac{1}{\Gamma(c)}$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} w^{a-1} (1-w)^{b-1} \cdot \frac{g(t)}{\Gamma(a+b) \cdot (1t)^{a+b} e^{-\lambda t} \cdot \frac{1}{\Gamma(c)}}$$

Story: Bank–post Office

③ T, w independent.

$$T \sim \text{Gamma}(a+b, \lambda)$$

$$w \sim \text{Beta}(a, b)$$

—

$$\frac{\pi(a+b)}{\pi(a)\pi(b)} w^{a-1} (1-w)^{b-1}$$

$$\frac{1}{\beta(a,b)} w^{a-1} (1-w)^{b-1}$$

④ \Rightarrow

$$\beta(a, b) = \frac{\pi(a)\pi(b)}{\pi(a+b)}$$



Story: Bank-post Office 借助 Γ 分布, 得出 β 分布期望

⑤ . $W \sim \text{Beta}(a, b)$, $a > 0, b > 0$. $E[W]$?

$$T = \underbrace{X+Y}_{\text{,}} \quad W = \frac{X}{X+Y} \quad X \sim \overbrace{\text{Gamma}(a, \lambda)}^{\text{,}}, \quad E(X) = \frac{a}{\lambda}$$
$$Y \sim \overbrace{\text{Gamma}(b, \lambda)}^{\text{,}} \quad E(Y) = \frac{b}{\lambda}$$

T, W are independent.

$$E[T \cdot W] = E[T] \cdot E[W]$$

$$\Rightarrow E[W] = \frac{E[T \cdot W]}{E[T]} = \frac{E[X]}{E[X+Y]} = \frac{E[X]}{E[X]+E[Y]}$$

$$= \frac{\frac{a}{\lambda}}{\frac{a}{\lambda} + \frac{b}{\lambda}} = \frac{a}{a+b}, \quad a > 0, b > 0$$

Outline

- 1 Change of Variables
- 2 Convolutions
- 3 Order Statistics
- 4 Beta-Binomial Conjugacy
- 5 Gamma Distribution
- 6 Dirichlet-Multinomial Conjugacy
- 7 Application Case: Bayesian Ranking

Story of Multinomial Distribution

Each of n objects is independently placed into one of k categories. An object is placed into category j with probability $\underline{p_j}$, where the p_j are nonnegative and $\sum_{j=1}^k p_j = 1$. Let X_1 be the number of objects in category 1, X_2 the number of objects in category 2, etc., so that $X_1 + \dots + X_k = n$. Then $\mathbf{X} = (X_1, \dots, X_k)$ is said to have the Multinomial distribution with parameters n and $\mathbf{p} = (\underline{p_1}, \dots, \underline{p_k})$. We write this as $\mathbf{X} \sim Mult_k(n, \mathbf{p})$.

贝叶斯统计学
多项式分布
dirichlet分布

Multinomial Joint PMF

Theorem

If $\mathbf{X} \sim \text{Mult}_k(n, \mathbf{p})$, then the joint PMF of \mathbf{X} is

$$P(X_1 = n_1, \dots, X_k = n_k) = \frac{n!}{n_1! n_2! \dots n_k!} \cdot p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$$

for n_1, \dots, n_k satisfying $n_1 + \dots + n_k = n$.

Dirichlet Distribution

$k=2$

↑
Beta 分布的推广

$$p_1^{\alpha_{1-1}} \cdot p_2^{\alpha_{2-1}}$$

$$(p_1^{\alpha_{1-1}} \cdot (1-p_1)^{\alpha_{2-1}})$$

The Dirichlet distribution is parameterized by a vector α of positive real numbers. The PDF is:

$$f(p_1, p_2, \dots, p_k; \alpha_1, \alpha_2, \dots, \alpha_K) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k p_i^{\alpha_{i-1}}$$

where $p_1 + \dots + p_k = 1$ and $0 < p_i < 1$.

C

Story: Dirichlet-Multinomial Conjugacy

① Unknown $P = (p_1, \dots, p_k)$. Prior distribution: Dirichlet

$$P = (p_1, \dots, p_k).$$

$$\text{PDF } f(p) = C \cdot p_1^{d_1-1} \cdot p_2^{d_2-1} \cdots p_k^{d_k-1}.$$

distribution.

② Data model: $X = (X_1, \dots, X_k)$ $X_1 + \dots + X_k = n$

$$X | P = p \sim \text{Multi}_k(n, p).$$

Define $\vec{n} = (n_1, \dots, n_k)$. $(X = \vec{n} \Leftrightarrow X_1 = n_1, X_2 = n_2, \dots, X_k = n_k)$

$$③ f(p | X = \vec{n}) = \frac{\Pr(X = \vec{n} | p)}{\Pr(X = \vec{n})} \cdot f(p) = \frac{n!}{n_1! \cdots n_k!} \cdot p_1^{n_1} \cdots p_k^{n_k} \cdot C \cdot p_1^{d_1-1} \cdots p_k^{d_k-1}$$

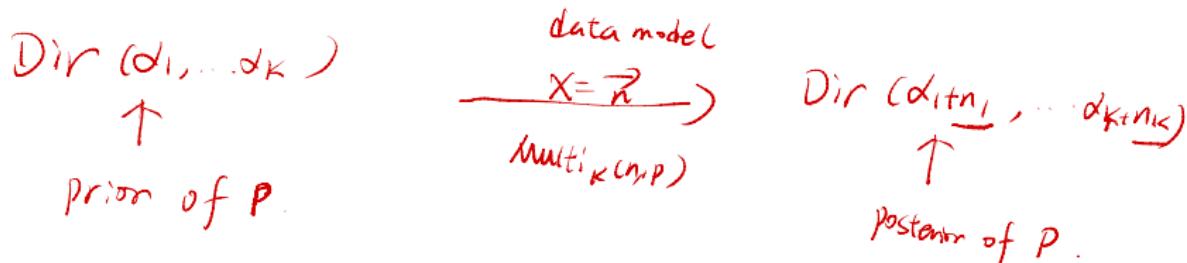
$$\propto p_1^{n_1+d_1-1} \cdots p_k^{n_k+d_k-1}$$

Still Dirichlet distribution.

$$\Pr(X = \vec{n})$$

↑ 与 p 无关,
看作常数

Story: Dirichlet-Multinomial Conjugacy



$\alpha_1, \dots, \alpha_K$

Pseudo count

$\alpha_i \rightarrow \alpha_i + n_i$

⋮

$\alpha_K \rightarrow \alpha_K + n_K$

Story: Dirichlet-Multinomial Conjugacy

If we have a Dirichlet prior distribution on \mathbf{p} and data that are conditionally Multinomial given \mathbf{p} , then when going from prior to posterior, we don't leave the family of Dirichlet distributions. We say that the Dirichlet is the conjugate prior of the Multinomial.

Outline

- 1 Change of Variables
- 2 Convolutions
- 3 Order Statistics
- 4 Beta-Binomial Conjugacy
- 5 Gamma Distribution
- 6 Dirichlet-Multinomial Conjugacy
- 7 Application Case: Bayesian Ranking

Rating System

- Consumers rely on the collective intelligence of other consumers: rating
- A common metric: 5 star rating 前提：样本数足够多
- Requirement: many ratings are needed to make this system work
- Quality of rating system depends on
 - ▶ average number of stars
 - ▶ average number of reviews

Which One to Choose?

- 1. Presto Coffee Pot - average rating of 5 (1 review).
- 2. Cuisinart Brew Central - average rating of 4.1 (78 reviews)

Example: Movie Ranking

- Data Set : <http://grouplens.org/datasets/movielens/>
- Top Ten Movies

Top 10 Movies chosen by Mean

算术平均

title	count	mean
Aiqing wansui (1994)	1	5
They Made Me a Criminal (1939)	1	5
Great Day in Harlem, A (1994)	1	5
Saint of Fort Washington, The (1993)	2	5
Entertaining Angels: The Dorothy Day Story (1996)	1	5
Someone Else's America (1995)	1	5
Star Kid (1997)	3	5
Santa with Muscles (1996)	2	5
Prefontaine (1997)	3	5
Marlene Dietrich: Shadow and Light (1996)	1	5

Tool: Bayesian Estimation

贝叶斯方法

- Mean of star reviews with a limited number of observations
- Useful for recommender services and other predictive algorithms that use preference space measures like star reviews.

Joint Distribution

- To use Bayesian estimation to compute the posterior probability for star ratings, we must use a joint distribution.
- We are not estimating the distribution of some scalar value X but, rather, the joint distributions of the probability estimate of whether or not the reviewer will give the movie a 1, 2, 3, 4, or 5 star rating (not just a simple thumbs up or down).
- In this case, the random variable is a categorical distribution because it can take some value within 1,2,3,4,5 with probabilities as follows:

$$p_1 + p_2 + p_3 + p_4 + p_5 = 1$$

Multinomial Distribution

- We can compute our posterior probability with N observations for five categories with corresponding numbers K_1, K_2, K_3, K_4, K_5 as follows:

$$Pr(O|p_1, p_2, p_3, p_4, p_5) \propto p_1^{K_1} p_2^{K_2} p_3^{K_3} p_4^{K_4} p_5^{K_5}$$

where $K_1 + \dots + K_5 = N$.

- This is a multinomial distribution.

Dirichlet Distribution: Prior

- If we include our prior as a distribution of the exact same form in the proportionality equation (e.g. a Dirichlet distribution with parameter α^0), then

$$Pr(p_1, p_2, p_3, p_4, p_5 | O) \propto \prod_{j=1}^5 p_j^{K_j + \alpha_j^0 - 1}$$

- This is another Dirichlet distribution with another parameter α^1 :

$$\alpha_j^1 = K_j + \underbrace{\alpha_j^0}_{\text{red underline}}, \forall j$$

Expected Average

- What is the expected value of the average rating given a posterior in the shape of our Dirichlet distribution?
- The expected value of the average rating based on the posterior is then computed for our star ratings as follows:

$$E(\underline{p_1} + \underline{2p_2} + \underline{3p_3} + \underline{4p_4} + \underline{5p_5}|O) = \sum_{i=1}^5 iE(p_i|O)$$

- Using our Dirichlet distribution we can compute the probability of a star value given our observations as the ratio of the Dirichlet parameter for that star to the sum of the Dirichlet parameters:

$$E(p_i|O) = \frac{\alpha_i^1}{\sum_{j=1}^5 \alpha_j^1}$$

Intra-Item: Bayesian Average Rating

$$\text{Bayes Average Rating} = \frac{\sum_{i=1}^5 i\alpha_i^0 + \sum_{i=1}^5 iK_i}{N + \sum_{i=1}^5 \alpha_i^0}$$

打分数

- N : the number of reviews $= \sum_{i=1}^5 k_i$
- $\sum_{i=1}^5 iK_i$: sum of all review scores
- $\sum_{i=1}^5 \alpha_i^0$: prior(given) number of reviews
- $\sum_{i=1}^5 i\alpha_i^0$: prior sum of all review scores

$$C = \sum_{i=1}^5 \alpha_i^0$$

估计数

估评分总和

Intra-Item: Bayesian Average Rating

α : average

$$1^{\circ} \text{. } N=0 ; \text{ Bayes Average} = \frac{C \cdot m}{C} = m. \quad \sum(\text{ratings}) = N \cdot \alpha$$

$$2^{\circ} \text{. } C=0 ; \text{ Bayes Average} = \frac{N \cdot \alpha}{N} = \underline{\alpha}$$

$$\text{Bayes Average Rating} = \frac{C \cdot m + \sum(\text{ratings})}{C + N}$$

- N : the number of reviews
- m : a prior for the average of review scores
- C : a prior for the number of reviews

加权平均 $= \frac{C \cdot m + N \cdot \alpha}{C + N}$

没有先验 Bayes 平均 = 算术平均

Example: Movie Ranking

- Data Set : <http://grouplens.org/datasets/movielens/>
- Top Ten Movies

Case 1: $m = \underline{3.25}$ & $C = \underline{50}$

title	bayes	count	mean
One Flew Over the Cuckoo's Nest (1975)	4.125796	264	4.291667
Raiders of the Lost Ark (1981)	4.145745	420	4.252381
Rear Window (1954)	4.167954	209	4.387560
The Silence of the Lambs (1991)	4.171591	390	4.289744
The Godfather (1972)	4.171706	413	4.283293
The Usual Suspects (1995)	4.206625	267	4.385768
Casablanca (1942)	4.250853	243	4.456790
The Shawshank Redemption (1994)	4.265766	283	4.445230
Star Wars (1977)	4.270932	583	4.358491
Schindler's List (1993)	4.291667	298	4.466443

Case 2: $m = \underline{2}$ & $C = \underline{6}$

拉低先验平均

title	count	bayes	mean
One Flew Over the Cuckoo's Nest (1975)	264	4.244526	4.291667
The Godfather (1972)	413	4.252955	4.283293
The Silence of the Lambs (1991)	390	4.257500	4.289744
Star Wars (1977)	583	4.335582	4.358491
The Usual Suspects (1995)	267	4.335740	4.385768
The Wrong Trousers (1993)	118	4.351562	4.466102
A Close Shave (1995)	112	4.368852	4.491071
The Shawshank Redemption (1994)	283	4.395904	4.445230
Casablanca (1942)	243	4.399209	4.456790
Schindler's List (1993)	298	4.418831	4.466443

Inter-Items: Pseudo Bayesian Average Rating

根据目录中的

$$\bar{m}_i = \frac{C_i \cdot m_i + \sum(\text{ratings})}{C_i + N}$$

$\sum(\text{ratings})$
 N prior

- \bar{m}_i : bayesian average rating for item i
- N : the number of reviews for all items
- m_i : average of review scores for item i
- C_i : the number of reviews for item i

Example: Bayesian Changes Order

$$\bar{m}_1 = \frac{c_1 \cdot m_1 + (\sum \text{ratings})}{c_1 + N} = \frac{(10 \times 4.92 + 2332.63)}{10 + 527} = 4.836$$

打分人数

MacBook	No. Ratings	Ave. Rating	Rank	Bayesian Rating	Bayesian Rank
MB991LL	10	4.920	1	4.436	2
MB403LL	15	4.667	2	4.433	3
MB402LL	228	4.535	3	4.459	1
MC204LL	150	4.310	4	4.401	5
MB061LL	124	4.298	5	4.402	4

$$N = 10 + 15 + 228 + 150 + 124 = 527;$$

$$\begin{aligned}\sum(\text{ratings}) &= 10 \times 4.92 + (15 \times 4.66) + 228 \times 4.535 + 150 \times 4.310 \\ &\quad + 124 \times 4.298 = 2332.637\end{aligned}$$

Reverse Engineering Amazon

- Bayesian adjustment
- Recency of view
- Reputation score
 - ▶ www.amazon.com/review/top-reviewers-classic

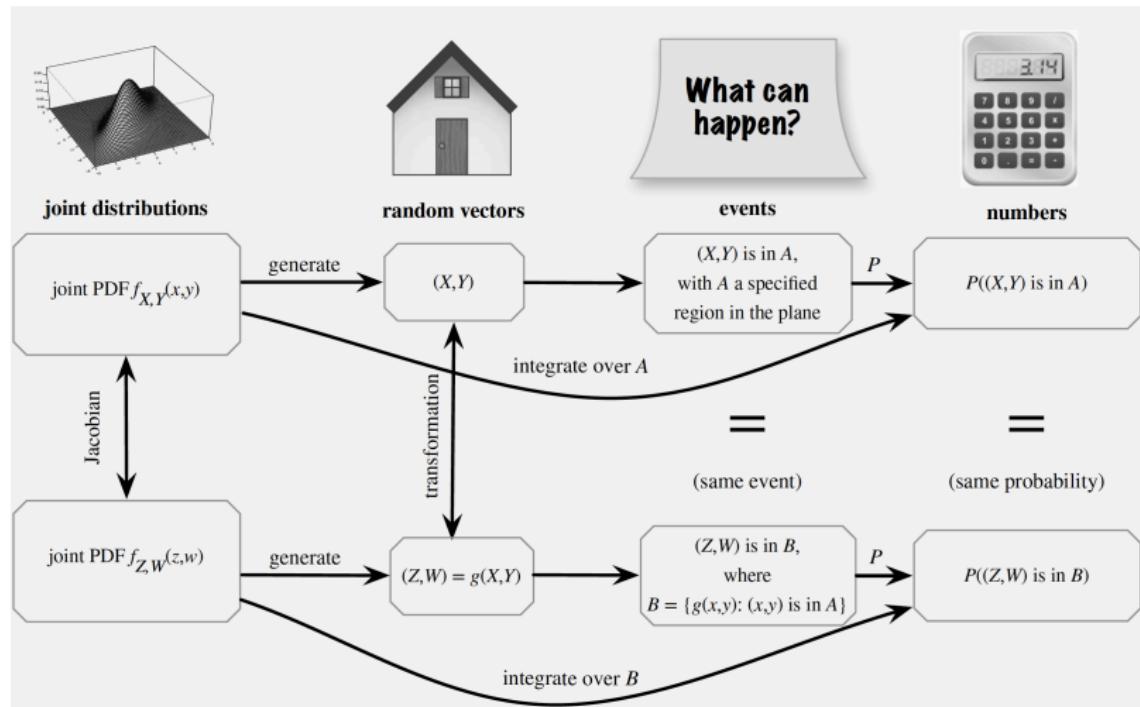
Key Factors

- Bayesian ranking
- Too few or too outdated reviews penalized
- Very high quality reviews help a lot

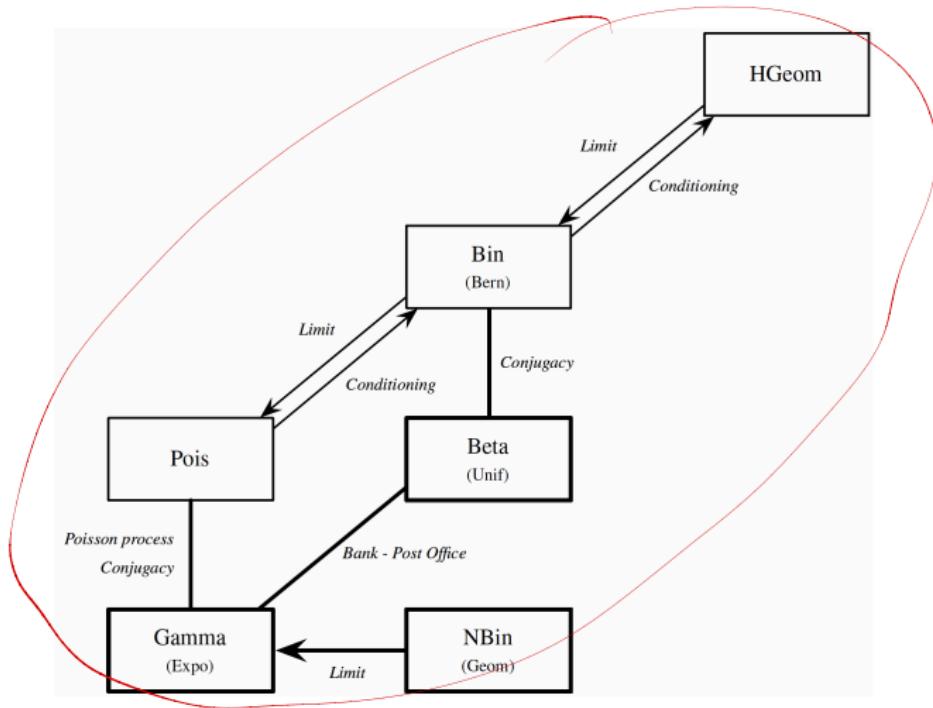
Summary

- Average ratings scalarize a vector and ranks
- Number of ratings should matter, Bayesian ranking does that
- Other statistical methods help too

Summary 1: Transformation



Summary 2



References

- Chapter 8 of **BH**
- Chapter 4 of **BT**