

Reminder:

cts R.V. pdf. $p(x)$ " $P\{X \in x \pm \frac{dx}{2}\} = p(x)dx$ "

other prob concepts:

expectation/mean of a R.V.

$E[X] \triangleq \sum_{x \in \Omega_X} x p(x)$ (for discrete R.V.)
 $\int_{\Omega} x p(x) dx$ ("cf. R.V.")

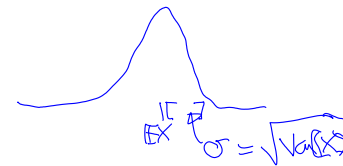
($E[X]$ does not always exist \rightarrow e.g. Cauchy distribution has integral diverging)

$\mathbb{E}[\cdot]$ is a linear operator $\Rightarrow \mathbb{E}[aX + Y] = a\mathbb{E}[X] + \mathbb{E}[Y]$

variance

$$\text{Var}[X] \triangleq \mathbb{E}[(X - \mathbb{E}X)^2] \\ = \mathbb{E}X^2 - (\mathbb{E}X)^2$$

→ "measure of dispersion"



independence:

X is independent of $Y \Leftrightarrow p(x,y) = p(x)p(y) \quad \forall x \in \Omega_x$
 $\forall y \in \Omega_y$
 notation: $X \perp Y$

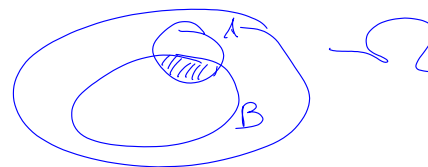
"mutually independent" for R.V. $X_1, \dots, X_n \Leftrightarrow p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p(x_i)$

conditioning:

• for events $A \subseteq B$, suppose that $P(B) \neq 0$

then define $P(A|B) \triangleq \frac{P(A \cap B)}{P(B)}$

we know that $P(B) = \sum_{A \in \text{partition of } \Omega} p(A \cap B)$



'conditional pmf'



• for a discrete R.V., define $P(X=x|Y=y) \triangleq \frac{p(X=x, Y=y)}{P(Y=y)}$ joint pmf
normalization constant { $P(Y=y)$ marginal pmf

same thing with "conditional pdf"

subtle point: $p(x|y)$ is undefined when $p(y)=0$

example: "P(having cancer | tumor measurement = a)"

⊗ Bayes rule → inverting the conditioning

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

"Bayes rule"

$$\text{by defn: } p(y|x) = \frac{p(x,y)}{p(x)}$$

$$\Rightarrow p(x,y) = p(y|x)p(x)$$

"product rule"

chain rule → successive application of product rule

$$\text{recall: } (x_{1:j} \triangleq (x_1, \dots, x_j))$$

$$p(x_1, \dots, x_n) = p(x_{1:n-1}) p(x_n | x_{1:n-1})$$

$$= \dots$$

always true? $= \prod_{i=1}^n p(x_i | x_{1:i-1})$
 can be simplified using cond. indep. assumptions
 [directed graph model]

cond. indep.: X is cond. indep. of Y given Z

notation: $X \perp\!\!\!\perp Y \mid Z$

$$\Leftrightarrow p(x, y | z) = p(x | z) p(y | z) \quad \begin{array}{l} \forall x \in \Omega_x \\ y \in \Omega_y \\ z \in \Omega_z \\ \text{s.t. } p(z) \neq 0 \end{array}$$

example: Z = indicator whether mother carries a genetic disease

X = son 1 has disease
 Y = son 2 has disease

$$X \perp\!\!\!\perp Y \mid Z$$

here, X is not "marginally indep." of Y i.e. $X \not\perp\!\!\!\perp Y$
 but X is cond. indep. of Y given Z

exercise to the reader: prove that if $X \perp\!\!\!\perp Y \mid Z$

$$\text{then } p(x | y, z) = p(x | z)$$

$$[\text{also, } X \perp\!\!\!\perp Y \Rightarrow p(x | y) = p(x)]$$

Parametric models:

← family of distributions

$$\mathcal{P}_\Theta = \{ p(\cdot | \theta) \mid \theta \in \Theta \}$$

$\cup \{ \}$ $\underbrace{\quad}$ possible pmf/pdf depending on θ
 $\bar{\theta}$ set of valid parameters

from context = "support of distribution" i.e. Ω_X
 is usually fixed for all θ

abuse of notation: $\sum p(x; \theta) | \theta \in \Theta$

"better" notation $\sum p(x; \theta) | \theta \in \Theta$

examples: $\Omega_X = \{0, 1\}$ for a coin flip

$\Omega_X = [0, +\infty[$ for gamma distribution

notation: $X \sim \text{Bern}(\theta)$

"R.v. X is distributed as a $\text{Bern}(\theta)$ dist."

$p(x; \theta) = \text{Bern}(x; \theta)$
 θ parameter of dist.
 x variable for pmf

another example: $p(x; (\mu, \sigma^2)) = N(x; \mu, \sigma^2)$

Bernoulli: prob. of a coin flip

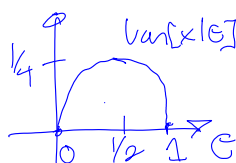
$$P\{X=1 | \theta\} = \theta$$

$$\Theta = [0, 1]$$

$$\Omega_X = \{0, 1\}$$

$$E[X] = \theta$$

$$\text{Var}[X] = \theta(1-\theta)$$



binomial distribution: models n independent coin flips

sum of n indep. $\text{Bern}(\theta)$ r.v.
 θ often denoted p

let $X_i \overset{\text{i.i.d.}}{\sim} \text{Bern}(\theta)$ independent & identically distributed

implicitly "is talking about" X_1, \dots, X_n

let $X = \sum_{i=1}^n X_i$; then we have $X \sim \text{Bin}(n, \theta)$ "binomial with..."

1=1

parameter $n \in \mathbb{N}$

$$\Omega_X = \{0, 1, \dots, n\}$$

$$p(x; \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} \text{ for } x \in \Omega_X$$

of ways
to get x successes
out of n trials

$$\binom{n}{x} \triangleq \frac{n!}{x!(n-x)!}$$

$$\theta^{\sum_{i=1}^n x_i} (1-\theta)^{\sum_{i=1}^n (1-x_i)} = \prod_{i=1}^n \text{Bern}(x_i; \theta)$$

mean $X = \sum_i X_i \in$

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X_i] = n\theta$$

similarly $\text{Var}[X] = n\theta(1-\theta)$

other distributions:

Poisson(λ)
 μ_{mean}

$$\Omega_X = \{0, 1, \dots\} = \mathbb{N}$$

[count data]

Gaussian (μ, σ^2)

$$N(\mu, \sigma^2)$$

$$\Omega_X = \mathbb{R}$$

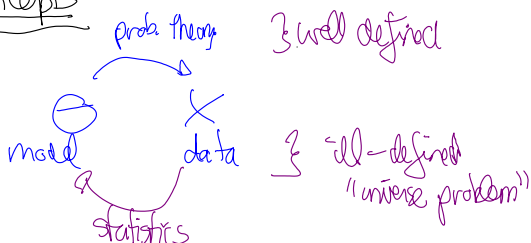
gamma $\Gamma(\alpha, \beta)$
 μ, μ

$$\Omega_X = \mathbb{R}_+$$

other: Laplace, Cauchy, exponential, beta, Dirichlet, etc.

Statistical concepts

cartoon



epistemology
of science

Observation: model is in nature. confirmation

example: model is n indep. coin flips

prob. theory: probs of k heads in a row

Statistics: I have observed k heads
 $n-k$ tails what is θ ?

Frequentist vs Bayesian:

philos. meaning of a probability?

a) (traditional) frequentist semantic

$P(X=x)$ represents the ~~limiting~~ relative frequency of observing $X=x$

\Leftrightarrow I could repeat to # of iid experiments

b) Bayesian (subjective) semantic

$P(X=x)$ encodes an agent "belief" that $X=x$

laws of probability characterizes a "rational" way to combine "beliefs"

and "evidence" [observations]

[\rightarrow has motivation in terms of gambling, utility/decision theory, etc...]

operationally:

(frequentist justification) • for a discrete R.V., suppose that $P\{X=x\} = \theta$

$$\Rightarrow P\{X \neq x\} = 1 - \theta$$

$B \triangleq \mathbb{1}\{X=x\} \sim \text{Bern}(\theta)$ R.V.

repeat iid. experiments i.e. $B_i \stackrel{\text{iid.}}{\sim} \text{Bern}(\theta)$

by L.L.N.
(law of large numbers)

$$\frac{1}{n} \sum_i B_i \xrightarrow{\text{a.s.}} E[B_1] = \theta$$

\nwarrow limiting frequency

[aside: $\sum_{i=1}^n B_i \sim \text{Bin}(n, \theta)$ $\mathbb{E}(\sum_{i=1}^n B_i) = n\theta = \epsilon$]

interpretation

$$\text{Var}[\sum_{i=1}^n B_i] = \frac{1}{n^2} \text{Var}[\text{Bin}(n, \theta)] = \frac{1}{n^2} n\theta(1-\theta) = \frac{1}{n} \theta(1-\theta)$$

(use $\text{Var}[aX] = a^2 \text{Var}[X]$)

$$\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n B_i - \theta \right) \xrightarrow{d} N(0, \theta(1-\theta))$$

CLT (central limit theorem)

→ all works

Bayesian approach:

⊛ very simple philosophically: treat all uncertain quantities as P.V.

→ i.e. encode all knowledge about the system ("beliefs")

as a "prior" on probabilistic models

and then use laws of prob. (and Bayes rule) to get answers?

simplest example: biased coin

we believe $X \sim \text{Bin}(n, \theta)$

unknown? \Rightarrow model it as a P.V.

\Rightarrow we need a $p(\theta)$ "prior distribution"

$$\mathcal{I}_{\theta} = [0, 1]$$

suppose we observe $X=x$ (result of n flips)

then, we can "update" our belief about θ using Bayes rule

$$p(\theta | x) \propto p(x | \theta) p(\theta)$$

prior belief

$$\underbrace{p(\theta|x) \sim \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta) d\theta}}_{\text{posterior belief}} \quad \left[\begin{array}{l} \text{normalization} \\ \text{distribution model / likelihood} \end{array} \right] \quad \text{"marginal likelihood"}$$

[note: $p(x|\theta) \rightarrow$ is pmf $p(x, \theta) \leftarrow$ is a mixed distribution]
 $p(\theta) \rightarrow$ pdf

example: suppose $p(\theta)$ is uniform on $[0,1]$ "no specific preference"

$$p(\theta|x) \propto \theta^x (1-\theta)^{n-x} \underbrace{\mathbb{1}_{[0,1]}(\theta)}_{p(\theta)} \quad (x \in 0:n)$$

"proportional to"

scaling: $\int_0^1 \theta^x (1-\theta)^{n-x} d\theta = B(x+1, n-x+1)$

$$\underbrace{B(a,b)}_{\text{Beta function}} \triangleq \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad \Gamma(a) \triangleq \int_0^\infty u^{a-1} e^{-u} du$$

$p(\theta|x)$ is called a "beta distribution"

$$B(\theta; \underbrace{\alpha, \beta}_{\text{parameters}}) \triangleq \frac{\theta^{\alpha-1} (1-\theta)^{\beta-1}}{B(\alpha, \beta)} \underbrace{\mathbb{1}_{[0,1]}(\theta)}_{\text{support}}$$