Probability Review Topics

1D distributions
  discrete (pmf) vs continuous (pdf)
  normalized vs unnormalized
  examples [1 2 1] ; uniform(0,1);   1-x^2 | -1 <= x <= 1;   N(0,1)

2D (bivariate) distributions
  joint distribution (with examples of discrete and continuous)
  examples [0 5 5; 10 0 0; 2 4 6];
           [1 1 1; 1 1 1; 1 1 1];
           [1 2 1; 2 4 2; 1 2 1];
           bivariate Gaussian
  marginal distributions
  conditional distributions
  independence
  conditional independence

Multivariate distributions
  (vector of random variables X=[x1 x2 x3 ... xn]

Cumulative Distribution Function (cdf)
  [note: tie in with integral images!]

Expectation / Expected Values
  moments and central moments
  mean, variance, covariance
  "one-pass" computation of central moments
    (useful for sequences / time series)
  moments of binary images = shape descriptors

Probability distribution functions

Let $X = \{x_1, x_2, \ldots, x_n\}$ discrete set

$f_X(x_i)$ is a pmf (prob mass function)

if
$$f(x_i) \geq 0$$
$$\sum_{i=1}^{n} f(x_i) = 1$$
$$P(X = x_i) = f(x_i)$$

Let $X = \mathbb{R}$ continuous

$f_X(x)$ is a pdf (prob density function)

if
$$f(x) \geq 0$$
$$\int_{-\infty}^{\infty} f(x) = 1$$
$$P(a \leq x \leq b) = \int_a^b f(x)\,dx$$

note: $P(X = x_i) = 0$ for continuous variables!

Important concept: normalized vs unnormalized

$$f(x) \geq 0$$
$$\int_{-\infty}^{\infty} f(x)\,dx < \infty$$

can be treated as an unnormalized distribution

we can turn it into a pdf by dividing by appropriate normalizing constant $C = \int_{-\infty}^{\infty} f(x)\,dx$

$$P(x) = f(x) / \int_{-\infty}^{\infty} f(x)$$

- often we learn distributions in unnormalized form (e.g. histograms)
- ~~some methods of inference can use unnormalized distributions directly~~ ~~do not need~~
- a lot of sound & fury in statistical estimation has to do with computing the normalizing constant for high-dimensional distributions
- however some methods of inference & sampling can use unnormalized distributions directly!

Bivariate distribution

pmf $f(x,y) = P(X=x, Y=y)$ = prob $X=x$ and $Y=y$

so-called Joint distribution (α)

pdf $f(x,y)$: $P((x,y) \in \overset{\text{some region of } \mathbb{R}}{A}) = \iint_A f(x,y)\,dx\,dy$

running example pmfs

$$\begin{bmatrix} 0 & 5 & 5 \\ 10 & 0 & 0 \\ 2 & 4 & 6 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

Marginal distributions.

Let $f(x,y)$ be a joint distribution

$$f_X(x) = P(X=x) = \sum_Y P(X=x, Y=y) = \sum_Y f(x,y) \quad \text{or } \int_Y f(x,y)\,dy$$

similarly

$$f_Y(y) = P(Y=y) = \sum_X P(X=x, Y=y) = \sum_X f(x,y) \quad \int_X f(x,y)\,dx$$

simplified notation (That helps to derive algebraic manipulations)

$$P(x) = \sum_Y P(x,y) \qquad P(y) = \sum_X P(x,y)$$

(α) Note! every meaningful statistical question can be answered (computed) from the joint distribution

Conditional distribution

$$P(x|y) = \frac{P(x,y)}{P(y)} = \frac{P(x,y)}{\sum_x P(x,y)}$$

also note: $P(x,y) = P(x|y)P(y) = P(y|x)P(x)$
example of factoring a joint distribution.

Statistical Independence

r.v.'s x and y are independent iff

$$P(x,y) = P(x)P(y)$$

note, in that case

$$P(x|y) = \frac{P(x,y)}{P(y)} = \frac{P(x)P(y)}{P(y)} = P(x)$$

Conditional Independence
a more subtle form of independence

$$P(x,y|z) = P(x|z)P(y|z)$$

note: this does not imply nor is it implied by
statistical independence

cdf (cumulative distribution function)

$$P(X \le x) \quad \longrightarrow \quad \begin{cases} \sum_{\tau=0}^{x} P(X=\tau) & \text{discrete } p \\ \int_{\tau=-\infty}^{x} p(\tau)\,d\tau & \text{continuous } p \end{cases}$$

$$P(X \le x,\ Y \le y)$$

note: The cdf is a continuous function (defined over $\mathbb{R}$) even for discrete probability distributions

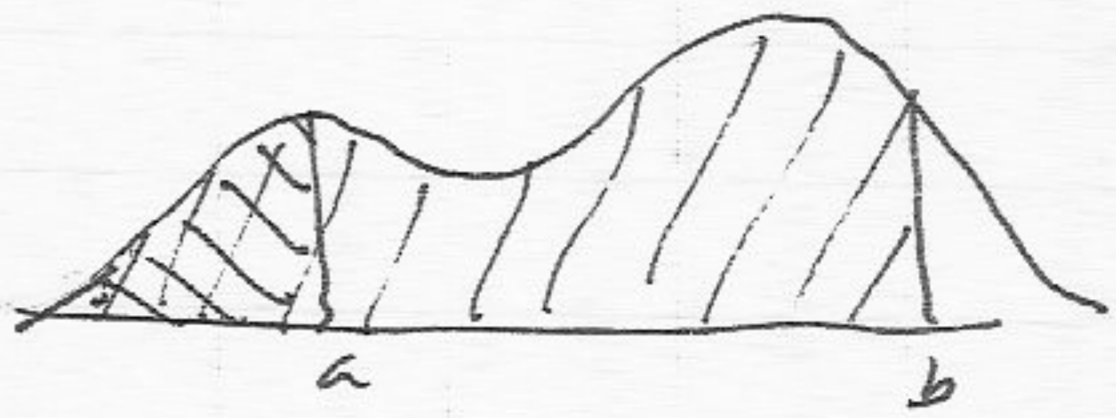
example

$$\begin{array}{cccc} & x=0 & x=1 & x=2 \\ f_X(x) = & \left[\ \tfrac{1}{4}\right. & \tfrac{1}{2} & \left.\tfrac{1}{4}\ \right] \end{array}$$

$$\text{cdf}(x) = F(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \le x < 1 \\ 3/4 & 1 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

note:

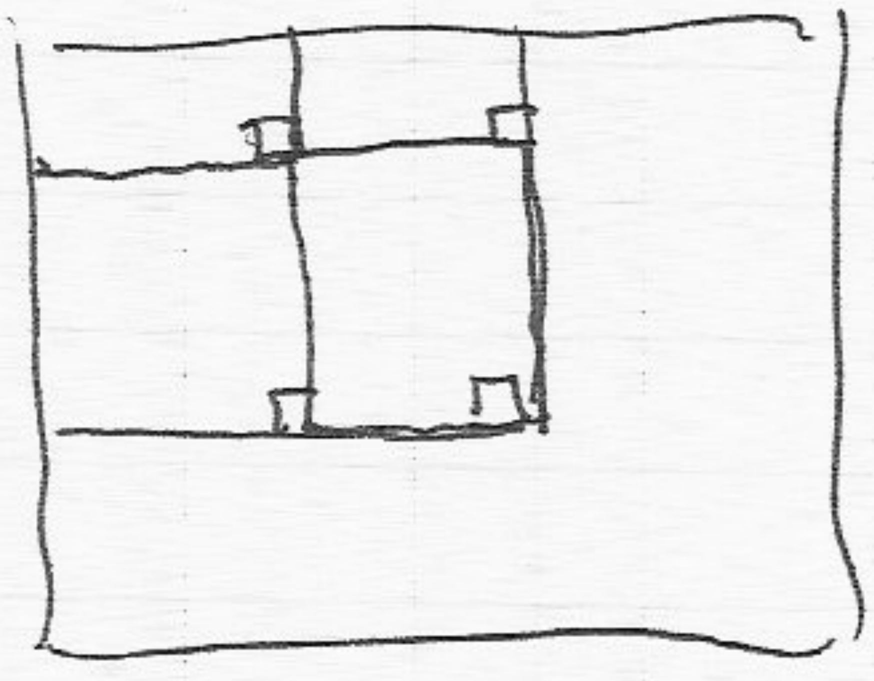
$$P(a \le x \le b) = F(b) - F(a)$$
$$= P(x \le b) - P(x \le a)$$



Tie in with integral images in vision

$$F(x,y) = \sum_{x=1}^{x} \sum_{y=1}^{y} I(x,y)$$

$$P(x_L \le x \le x_H,\ y_L \le y \le y_H)$$



$$F(x_H, y_H) - F(x_H, y_L - 1)$$
$$- F(x_H - 1, y_L) + F(x_L - 1, y_L - 1)$$

Expected values

$$E_x(g(x)) = \int_{x=-\infty}^{\infty} g(x) f(x)\,dx \quad \longleftarrow \text{expected value}$$

examples

"mean" $$E_x(x) = \int_{x=-\infty}^{\infty} x f(x)\,dx$$ also known as 1st moment

discrete example of mean $\quad P(x) = \begin{array}{ccc} x=0 & x=1 & x=2 \\ \end{array}$ $\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} * \frac{1}{6}$

$$E(x) = 0 \cdot \frac{1}{6} + 1 \cdot \frac{2}{6} + 2 \cdot \frac{3}{6}$$

$$= \frac{1}{3} + 1 = \frac{4}{3}$$

Note that this is not a value (in this example) that can be taken by this discrete random variable!

mode of this example = argmax of largest probability = 2
it is the "most probable" value

2nd central moment vs variance

$$E(x^2) \quad \text{vs} \quad E((x-\mu)^2)$$

$$E[(x-\mu)^2] = E[x^2 - 2\mu x + \mu^2]$$

note: expectation is a linear operator!

$$= Ex^2 - 2\mu \underbrace{Ex}_{\mu} + \mu^2$$

$$= E[x^2] - (E[x])^2$$

a function of 1st & 2nd central moments

$$\text{cov}(x,y) = E_{xy}[(x-\mu_x)(y-\mu_y)]$$