"Last" time: · x is r.v.

Best estimate of x is $E[x]$, minimizer of $E[(x-c)^2]$

· x is r.v. y is another r.v.

Best estimate of x as a fcn of y: $E[x|y]$

$E[(x-E(x|y))^2] \leq E[(x-g(y))^2]$ for any fcn g.

· Conditional probability: $p(x|y) = \dfrac{p(x,y)}{p(y)}$

$\boxed{\text{Note } \int p(x|y)\,dx = 1 \text{ but } \int p(x|y)\,dy \neq 1}$

$$E[x|y] = \int_{-\infty}^{\infty} x\, p(x|y)\,dx$$

· Independence: x, y are independent $\Leftrightarrow$ $p(x,y) = p(x)p(y)$

If x, y are independent: $p(x|y) = \dfrac{p(x,y)}{p(y)} = p(x)$

→ occurrence of y does not affect probability of x

→ knowing y does not mean that you know more about x.

$E[x|y] = \int x\, p(x|y)\,dx = \int x\, p(x)\,dx = E[x]$

best estimate with y = best estimate without y

Bayes' rule:

$$p(x|y) = \frac{p(x,y)}{p(y)}$$
$$p(y|x) = \frac{p(x,y)}{p(x)}$$
$$\Big\} \quad p(x|y) = p(y|x)\, p(x) \cdot \frac{1}{p(y)}$$

Interpretation:

- x is an unknown quantity you care about.
- what you know about x is contained in "prior" $p(x)$

Example:
- If you know that $x \geq 0$ (density, permeability, ...) then $p(x) = \mathcal{U}[0, \infty]$
- If you know that x is likely in interval $[\mu - 2\sigma, \mu + 2\sigma]$, $p(x) = \mathcal{N}(\mu, \sigma^2)$

- y is another r.v., connected to x.
- y could be a measurement of x:

"data"

$$y = h(x) + \eta$$

measurement model

r.v.
how far off is the measurement from the actual value of y?

- you make assumptions about $\eta \sim p_\eta(\eta)$

$$y - h(x) = \eta$$

"likelihood"

$$\Rightarrow \quad p_\eta(\eta) = p_\eta(y - h(x)) = p(y|x)$$

Altogether:

$$p(x|y) = p(y|x) \quad p(x) \quad \cdot \quad \frac{1}{\ }$$

Example $\quad p(x) = \mathcal{N}(0, I)$

$\dim(x) = \dim(y) = n$

$y = x + \eta, \quad \eta \sim \mathcal{N}(0, I)$

$$p(x|y) \propto p(x)\,p(y|x) \propto \exp\left(-\tfrac{1}{2} x^T x\right) \exp\left(-\tfrac{1}{2}(x-y)^T(x-y)\right)$$

$$p(x|y) \propto \exp(-F(x)), \quad F(x) = \tfrac{1}{2} x^T x + \tfrac{1}{2}(x-y)^T(x-y)$$

HW: If $q(x) \propto \exp(-F(x))$, $F(x)$ is quadratic, then $q(x)$ is Gaussian $\mathcal{N}(\mu, H^{-1})$, $\mu = \arg\min F$ & H is Hessian of F at μ.

$$\nabla F = x + (x-y) = 0 \Rightarrow \mu = \tfrac{1}{2} y$$

$$H = \nabla^2 F = 2I.$$

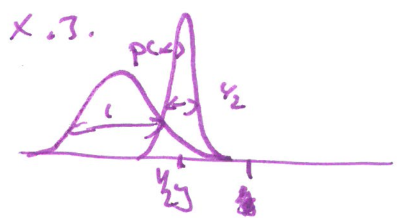
$$p(x|y) = \mathcal{N}\left(\tfrac{1}{2} y, \tfrac{1}{2} I\right)$$

updated, posterior mean

updated, posterior covariance

→ data "narrowed" down where we think x is.

→ 1D illus:



The main portion of this class is about how to ~~deal~~ compute with posterior distributions implicitly defined by differential equations.

Some useful tricks ~~we~~ we will use of ken:

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

2) $p(x,y|z) = p(x|y,z)\,p(y|z)$

Why? $p(x,y|z) = p(x,y,z)/p(z)$

$p(x|y,z) = p(x,y,z)/p(y,z)$

$$\left.\begin{array}{l} p(x,y|z) = p(x,y,z)/p(z) \\ p(x|y,z) = p(x,y,z)/p(y,z) \end{array}\right\} p(x,y|z) = p(x|y,z)\frac{p(y,z)}{p(z)}$$

$= p(x|y,z)\,p(y|z)$.

3) $p(x|y,z) = p(x|y)$ if z is independent of x,y.

Why? $p(x|y,z) = p(x,y|z)\dfrac{1}{p(y|z)} = \dfrac{p(x,y)}{p(y)} = p(x|y)$.

4) $x \sim N(\mu, P)$

$y = \underbrace{Ax+b}_{g(x)}$, A is invertible $\qquad y = N(A\mu+b, APA^T)$

Why? Change of vars from last time:

$$P_y(y) = P_x(g^{-1}(y))\left|\frac{dx}{dy}\right| \qquad x = A^{-1}(y-b)$$

$$\left|\frac{dx}{dy}\right| = \left|\det A^{-1}\right| = \frac{1}{\det A}$$

$$P_y(y) = (2\pi)^{-n/2}(\det P)^{-1/2}\exp\left(-\frac{1}{2}\left(A^{-1}(y-b)-\mu\right)^T P^{-1}\left(A^{-1}(y-b)-\mu\right)\right) \times \frac{1}{\det A}$$

Basic Monte Carlo

$$\eta = \frac{1}{n} \sum_{i=1}^{n} \eta_i$$

η_i are iid

(= independently identically distributed)

$E[\eta_i] = \bar{\eta}$

$Var(\eta_i) = \sigma^2$

η_i, η_j are independent if $i \neq j$

$P_i(\eta_i) = p(\cdot)$

Recall Chebyshev: $Pr(|x - E[x]| \geq k\sigma_x) \leq \frac{1}{k^2}$

$\sigma_x^2 = var(x)$.

Use this on η:

$$Pr(|\eta - E[\eta]| \geq k\sigma_\eta) \leq \frac{1}{k^2}$$

$$E[\eta] = E\left[\frac{1}{n}\sum \eta_i\right] = \frac{1}{n}\sum E[\eta_i] = \bar{\eta}$$

$$var(\eta) = var\left(\frac{1}{n}\sum \eta_i\right) = \frac{1}{n^2}\sum var(\eta_i) = \sigma^2/n$$

Chebyshev:

$$Pr\left(|\eta - \bar{\eta}| \geq \frac{k\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2}$$

Rules you probably know or should review:

$E[x+y] = E[x] + E[y]$

$var(\alpha x) = \alpha^2 var(x)$

$var(x+y) = var(x) + var(y)$ if x, y are independent

Message: for large n: $\eta \simeq \bar{\eta}$.

This work for more general expected values as well.

$$E[f(\eta)] \simeq \frac{1}{n} \sum f(\eta_i) = \bar{f}(\eta) \qquad \bar{f}(\eta) = \frac{1}{n} \sum f(\eta_i)$$

$$\Pr\left(|f(\eta) - \bar{f}(\eta)| \geq \frac{k\,\sigma(f(\eta))}{\sqrt{n}} \right) \leq \frac{1}{k^2}$$

↳ To compute an expected value of a r.v.:

$$E[f(x)] \simeq \frac{1}{n} \sum f(x_i), \qquad x_i \sim p(x)$$

- draw a large # of independent samples
- average.

Questions we will address in this class:

- What is "large"
  When is "large" large enough
- How does one draw samples from a given distribution?

Examples of this you have seen before:

1) $\bar{x} = E[x] \simeq \frac{1}{n} \sum_{i=1}^{n} x_i$

2) Covariance: $\mathrm{Cov}(x) = E[(x-\bar{x})(x-\bar{x})^T] \simeq \frac{1}{n} \sum (x_i - \bar{x})(x_i - \bar{x})^T$

What is $E[\hat{\sigma}^2] \overset{?}{=} \sigma^2$

x_i iid

$E[x_i] = \mu$

$var(x_i) = \sigma^2$

Check:

$$E[\hat{\sigma}^2] = E\left[\frac{1}{n} \sum_i (x_i - \bar{x})^2\right]$$

$$= \frac{1}{n}\left(\sum_i E[x_i^2] - 2E\left[x_i \frac{1}{n}\sum_j x_j\right] + E\left[\frac{1}{n^2}\left(\sum x_j\right)\left(\sum x_k\right)\right]\right)$$

$$= \frac{1}{n} \sum_i E[x_i^2] - \frac{2}{n} E[x_i^2] - \frac{2}{n} E\left[x_i \sum_{j \neq i} x_j\right]$$

$$+ \frac{1}{n^2}\left(\sum_l E[x_l^2] + E\left[\sum_k \sum_{j \neq k} x_k x_j\right]\right)$$

$$E\left[x_i \sum_{i \neq j} x_j\right] = E[x_i] \sum_{j \neq j} E[x_j] = \mu (n-1)\mu = (n-1)\mu^2$$

$$\sum_l E[x_l^2] + E\left[\sum_k \sum_{j \neq k} x_k x_j\right] = n\, E[x_i^2] + \mu^2 n(n-1)$$

$$\Rightarrow E[\hat{\sigma}^2] =$$

$$= \frac{1}{n} \sum_i E[x_i^2] - \frac{2}{n} E[x_i^2] - 2\frac{n-1}{n}\mu^2 + \frac{1}{n^2}\left(n\, E[x_i^2] + \mu^2 n(n-1)\right)$$

$$= \frac{1}{n} \sum_i E[x_i^2] \underbrace{\left(1 - \frac{2}{n} + \frac{1}{n}\right)}_{\frac{n-1}{n}} - \frac{n-1}{n}\mu^2$$

$\Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \hat{x})^2$ is not "unbiased".

$\frac{n}{n-1} \hat{\sigma}^2 = \frac{1}{n-1} \sum (x_i - \hat{x})^2$ is unbiased.

$\leadsto$ we use, for cov.: $\quad \mathrm{cov}(x) \simeq \frac{1}{n-1} \sum (x_i - \hat{x})(x_i - \hat{x})^T$

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

You can check that $\hat{x}$ is unbiased, i.e.

$$E[\hat{x}] = E[x_i] = \mu.$$