Probability Theory Review: 2-11

- common pdfs: Normal, Uniform, Exponential
- how does kernel density estimation work?
- common pmfs: Binomial (Bernoulli), Discrete Uniform, Geometric
- cdfs (and how to transform out from a random number generator (i.e. uniform distribution) into another distribution)
- how to plot: pdfs, cdfs, and pmfs in python.
- MLE revisited: how to derive the parameter estimate from the likehood function

Maximum Likelihood Estimation (parameter estimation)

Given data and a distribution, how does one choose the parameters?

likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

log-likelihood function:

$$l(\theta) = log \sum_{i=1}^{n} f(X_i; \theta)$$

maximum likelihood estimation: What is the θ that maximizes *L*?

Example: $X_1, X_2, \ldots, X_n$ ~ Bernoulli(p), then $f(x;p) = p^x(1 - p)^{1-x}$, *for* $x = 0, 1$.

$$L_n(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-X_i} = p^S(1-p)^{n-S}, \text{ where } S = \sum_i X_i$$

$$l_n(p) = S\, log\, p + (n - s)\, log(1 - p)$$

take the derivative and set to 0 to find: $\hat{p} = \dfrac{S}{n}$

Maximum Likelihood Estimation

Given data and a distribution, how does one choose the parameters?

likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

log-likelihood function:

$$l(\theta) = log \sum_{i=1}^{n} f(X_i; \theta)$$

maximum likelihood estimation: What is the θ that maximizes *L*?

Example: $X \sim$ Normal(μ, σ), then

GOAL: take the derivative and set to 0 to find:

Maximum Likelihood Estimation

Given data and a distribution, how does one choose the parameters?

likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

log-likelihood function:

$$l(\theta) = log \sum_{i=1}^{n} f(X_i; \theta)$$

maximum likelihood estimation: What is the θ that maximizes *L*?

Example: $X \sim$ Normal(μ, σ), then $f(x_1, x_2, \ldots, x_n; \mu, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$

$$\boxed{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

Normal pdf

GOAL: take the derivative and set to 0 to find:

Maximum Likelihood Estimation

Example: $X \sim$ Normal(μ, σ), then $$f(x_1, x_2, \ldots, x_n; \mu, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}}$$

GOAL: take the derivative and set to 0 to find:

Maximum Likelihood Estimation

Example: $X \sim$ Normal(μ, σ), then $f(x_1, x_2, ..., x_n; \mu, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}}$

$$log(f(x_1, ..., x_n; \mu, \sigma)) = -n \log \sqrt{2\pi} - n \log \sigma - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}$$

GOAL: take the derivative and set to 0 to find:

Maximum Likelihood Estimation

Example: $X \sim$ Normal(μ, σ), then $f(x_1, x_2, \ldots, x_n; \mu, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}}$

$$log(f(x_1, \ldots, x_n; \mu, \sigma)) = -n \log \sqrt{2\pi} - n \log \sigma - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}$$

first, we find μ using partial derivatives*:*

$$\frac{\partial f(x_1, \ldots, x_n; \mu, \sigma)}{\partial \mu} = \sum_{i=1}^{n} \frac{(x_i - \mu)}{2\sigma^2} = 0$$

GOAL: take the derivative and set to 0 to find:

Maximum Likelihood Estimation

Example: $X \sim$ Normal(μ, σ), then $f(x_1, x_2, \ldots, x_n; \mu, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}}$

$$log(f(x_1, \ldots, x_n; \mu, \sigma)) = -n \log \sqrt{2\pi} - n \log \sigma - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}$$

first, we find μ using partial derivatives:

$$\frac{\partial f(x_1, \ldots, x_n; \mu, \sigma)}{\partial \mu} = \sum_{i=1}^{n} \frac{(x_i - \mu)}{2\sigma^2} = 0 \;, \quad \sum_{i=1}^{n} x_i - n\mu = 0 \;, \quad \frac{\sum_{i=1}^{n} x_i}{n} = \hat{\mu} = \bar{x}$$

Maximum Likelihood Estimation

Example: $X \sim$ Normal(μ, σ), then $f(x_1, x_2, \ldots, x_n; \mu, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$

$$log(f(x_1, \ldots, x_n; \mu, \sigma)) = -n \log \sqrt{2\pi} - n \log \sigma - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}$$

first, we find μ using partial derivatives:

$$\frac{\partial f(x_1, \ldots, x_n; \mu, \sigma)}{\partial \mu} = \sum_{i=1}^{n} \frac{(x_i - \mu)}{2\sigma^2} = 0 \; , \quad \sum_{i=1}^{n} x_i - n\mu = 0 \; , \quad \frac{\sum_{i=1}^{n} x_i}{n} = \hat{\mu} = \bar{x}$$

now σ:

$$\frac{\partial f(x_1, \ldots, x_n; \mu, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^3} = 0$$

Maximum Likelihood Estimation

Example: $X \sim$ Normal(μ, σ), then $f(x_1, x_2, \ldots, x_n; \mu, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}}$

$$log(f(x_1, \ldots, x_n; \mu, \sigma)) = -n \log \sqrt{2\pi} - n \log \sigma - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^2}$$

first, we find μ using partial derivatives*:*

$$\frac{\partial f(x_1, \ldots, x_n; \mu, \sigma)}{\partial \mu} = \sum_{i=1}^{n} \frac{(x_i - \mu)}{2\sigma^2} = 0 \quad , \quad \sum_{i=1}^{n} x_i - n\mu = 0 \quad , \quad \frac{\sum_{i=1}^{n} x_i}{n} = \hat{\mu} = \bar{x}$$

now σ:

$$\frac{\partial f(x_1, \ldots, x_n; \mu, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^3} = 0 \quad , \quad \frac{n}{\sigma^2} + \sum_{i=1}^{n} (x_i - \mu)^2 = 0 \quad , \quad \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 = \hat{\sigma}^2$$

Maximum Likelihood Estimation

Example: $X \sim$ Normal(μ, σ), then $f(x_1, x_2, \ldots, x_n; \mu, \sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$

$$log(f(x_1, \ldots, x_n; \mu, \sigma)) = -n \log \sqrt{2\pi} - n \log \sigma - \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}$$

first, we find μ using partial derivatives:

$$\frac{\partial f(x_1, \ldots, x_n; \mu, \sigma)}{\partial \mu} = \sum_{i=1}^{n} \frac{(x_i-\mu)}{2\sigma^2} = 0 \;,\quad \sum_{i=1}^{n} x_i - n\mu = 0 \;,\quad \boxed{\frac{\sum_{i=1}^{n} x_i}{n} = \hat{\mu} = \bar{x}}$$

sample mean

now σ:

$$\frac{\partial f(x_1, \ldots, x_n; \mu, \sigma)}{\partial \sigma} = -\frac{n}{\sigma} + \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{\sigma^3} = 0 \;,\quad \frac{n}{\sigma^2} + \sum_{i=1}^{n} (x_i-\mu)^2 = 0 \;,\quad \boxed{\frac{1}{n}\sum_{i=1}^{n} (x_i-\mu)^2 = \hat{\sigma}^2}$$

sample variance

Maximum Likelihood Estimation

Try yourself:

Example: $X \sim \text{Exponential}(\lambda)$,

hint: should arrive at something almost familiar; then recall $\lambda = \dfrac{1}{\beta}$

Expectation, revisited

Conceptually: Just given the distribution and no other information: what value should I expect?

Expectation, revisited

Conceptually: Just given the distribution and no other information: what value should I expect?

Formally: The ***expected value*** of X is:

$$\mathbf{E}(X) = \int x\,dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $\mathbf{E}(X) = \mathbf{E}X = (x) = \mu = \mu x$

Expectation, revisited

Conceptually: Just given the distribution and no other information: what value should I expect?

Formally: The ***expected value*** of X is:

$$\mathbf{E}(X) = \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $\mathbf{E}(X) = \mathbf{E}X = (x) = \mu = \mu x$

"expectation" "mean" "first moment"

Expectation, revisited

Conceptually: Just given the distribution and no other information: what value should I expect?

Formally:

The ***expected value*** of X is:

$$\mathbf{E}(X) = \int x dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) dx & \text{if } X \text{ is continuous} \end{cases}$$

denoted: $\mathbf{E}(X) = \mathbf{E}X = (x) = \mu = \mu x$

“expectation” “mean” “first moment”

Alternative Conceptualization: If I had to summarize a distribution with only one number, what would do that best?

(the average of a large number of randomly generated numbers from the distribution)