TAM 598 Lecture 9:

MONTE CARLO ESTIMATES OF CDF, PDF, EXPECTATION, ETC

Announcements:

- HW 2 covers lectures 4-8; due on Feb 26

I. Cumulative Distribution Function

say $Y = RV$, $F(Y) = $ CDF.
we draw samples of Y and want to estimate $F(Y)$.

How do we do this? Note that we can write $F(y)$ as an expectation

$$F(y) = p(Y \leq y) = \int_{-\infty}^{y} p(y')\,dy' = \underbrace{\int_{-\infty}^{\infty} 1_{[-\infty,y]}(y')\,p(y')\,dy'}$$

$$= \mathbb{E}\left[1_{[-\infty,y]}(Y)\right] \quad \text{where } 1_{[-\infty,y]}(Y) = \begin{cases} 1 & \text{if } Y \in [-\infty, y] \\ 0 & \text{else} \end{cases}$$

"indicator function"

we can use our Monte Carlo estimator for $\mathbb{E}$:

(1) draw ind. samples $Y_1, Y_2, \ldots, Y_N$ from $p(y)$

(2) evaluate $\hat{F}_N(y) = \frac{1}{N}\sum_{i=1}^{N} 1_{[-\infty,y]}(Y_i) = \frac{\#\text{ of } Y_i \leq y}{N}$ ①

Alternatively, say we have X = RV, function $g(x)$ and say we want the CDF of $Y = g(x)$.

Then
$$F(y) = \mathbb{E}\left[\mathbb{1}_{[-\infty, y]}(Y)\right]$$
$$= \mathbb{E}\left[\mathbb{1}_{[-\infty, y]}(g(x))\right]$$

and again we use our Monte Carlo estimator for expectation:

(1) draw independent samples $X_1, X_2, \dots, X_N$ of X

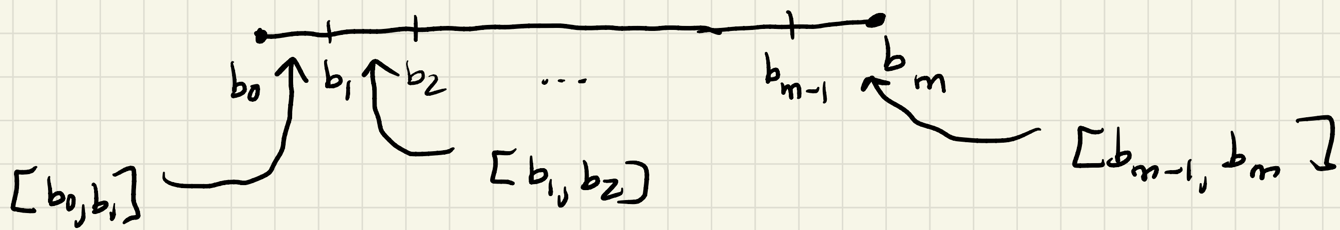
(2) $$\bar{F}_N(y) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{[-\infty, y]}(g(X_i))$$

$$= \frac{\#\text{ of } g(X_i) \leq y}{N}$$

II. Probability Density Function

- X = R.V.; $Y = g(X)$
- want to approximate the pdf $p(y)$ of $Y = g(X)$ given samples $X_1, X_2, \dots, X_N$

divide the domain of Y into small bins



we will approximating $p(y)$ by a piece wise constant function using its histogram

$$\hat{P}_M(y) = \sum_{j=1}^{M} c_j \, \mathbb{1}_{[b_{j-1}, b_j]}(y)$$

c_j are constants to be determined by sampling

The constant c_j is the probability that y falls in $[b_{j-1}, b_j]$

$$c_j = P(b_{j-1} \leq Y \leq b_j)$$

$$= F(b_j) - F(b_{j-1})$$

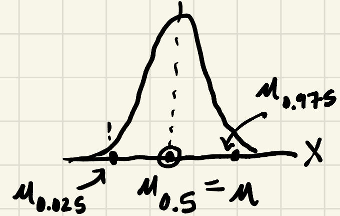
our estimate $\bar{c}_{j,N} = \bar{F}_N(b_j) - \bar{F}_N(b_{j-1})$

$$= \frac{\#\ \text{samples in bin } [b_{j-1}, b_j]}{N}$$

so $\hat{p}_{M,N}(y) = \sum_{j=1}^{M} \bar{c}_{j,N} \, \mathbb{1}_{[b_{j-1}, b_j]}(y)$

III. Estimating Quantiles u_q

$X = RV, \quad Y = g(X)$



$0 \le q \le 1$

$$F(u_q) = P(Y \le u_q) = q$$

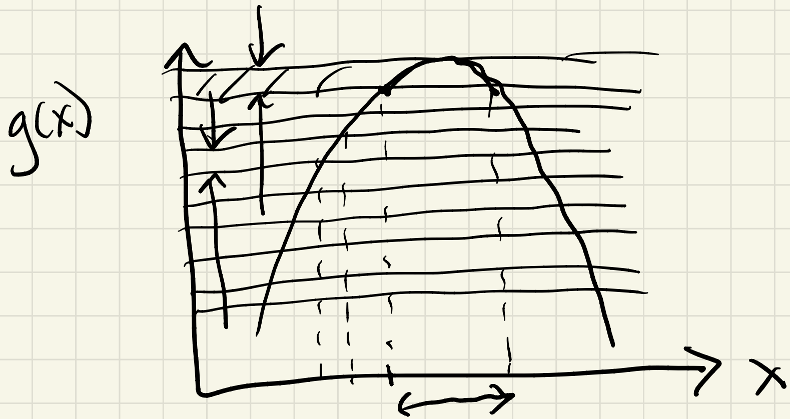
u_q is the q-quantile

approach: (1) Find the CDF

(2) solve $F(u_q) = q$ for u_q

$u_{0.5}$ = median, central value

$[u_{0.025}, u_{0.975}]$ = 95% predictive interval



g(x)
x

IV. Example – Propagating Uncertainty Through an ODE

$$\dot{y} = \frac{dy}{dt} = -a\,y(t)$$

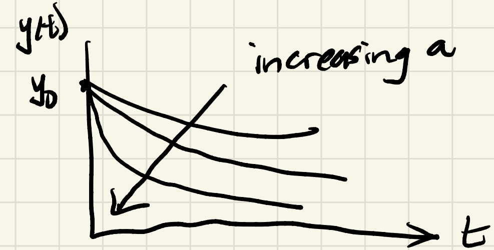
a = exponential decay constant, units t^{-1}

IV. Example - Propagating Uncertainty Through an ODE

$$\dot{y} = \frac{dy}{dt} = -a\,y(t)$$

a = exponential decay constant, units t^{-1}

initial conditions: $y(t=0) = y_0$



solution $y(t) = y_0 \exp(-at)$

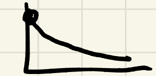
say we are uncertain about both the decay constant a and the initial conditions y_0

$\Rightarrow$ turn each uncertain quantity into a RV and sample

$\Rightarrow$ Need to select distributions for each RV

① Decay Rate a

we know $a > 0$ $\Rightarrow$ need a pdf with positive support

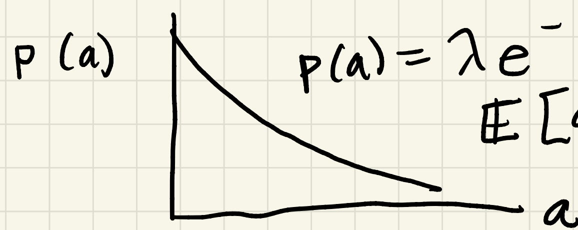
→ $\boxed{\text{Exponential}}$ 

→ Log-Normal

→ uniform $\mathcal{U}[0, b]$

Say we know $\mathbb{E}[a] = 0.1$, and we know nothing else

→ choose Exponential based on principle of max entropy



$$p(a) = \lambda e^{-\lambda a}$$

$$\mathbb{E}[a] = \frac{1}{\lambda}$$

choose $\lambda = 10$

$$p(a) = 10 e^{-10a}$$

② Initial Conditions y_0

we know $y_0 > 0$ so exp or L-N or uniform $\mathcal{U}[0,b]$ ...

say we know $\mathbb{E}[y_0] = 10$ and $\mathbb{V}[y_0] = 1$

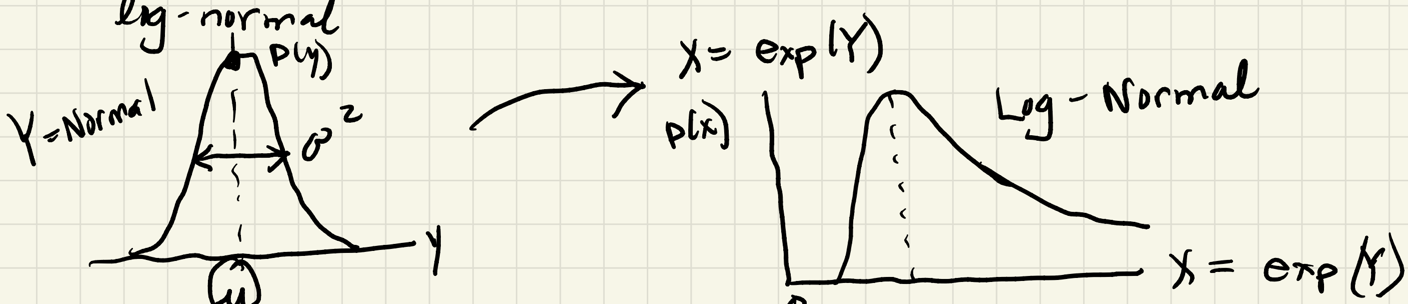
and we know nothing else.

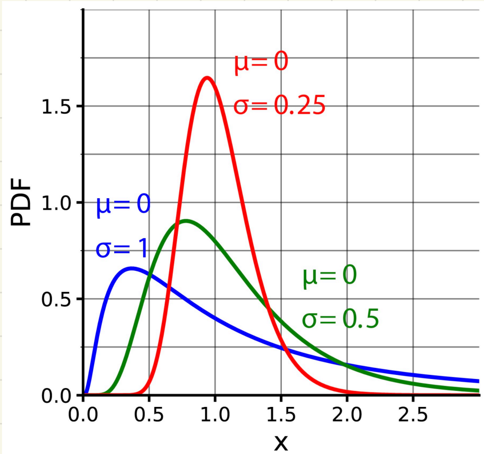
$\rightarrow$ choose $y_0 \sim \text{Log-Normal}(\mu, \sigma^2)$

Note: L-N is a continuous pdf whose logarithm is normally distributed.

If X is lognormal, then $Y \sim \ln(X)$ is normal

Equivalently if Y is gaussian, then $X \sim \exp(Y)$ is log-normal





lognormal (μ, σ^2)

pdf $\quad p(x) = \dfrac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\dfrac{\ln(x-\mu)^2}{2\sigma^2}\right)$

$$\begin{cases} \mathbb{E}[X] = \exp\left(\mu + \dfrac{\sigma^2}{2}\right) \\ \mathbb{V}[X] = \left|\exp(\sigma^2) - 1\right| \exp\left(2\mu + \sigma^2\right) \end{cases}$$

solving numerically: $\mu = 2.25$

$\sigma = 0.10$

UNCERTAINTY PROPAGATION:

Now we are ready to draw samples from a and y_0, and evaluate the statistics associated w/ our ODE

Note: y_0, a are RV's so $y(t)$ for any t is random

We call $y(t)$ a **random process** because it is parametrized by a label (time t)

our samples will 2D lists "y_samples"

time t

samples a, y_0	t_0	t_1	....	t_K
1				
2				
⋮				
N				