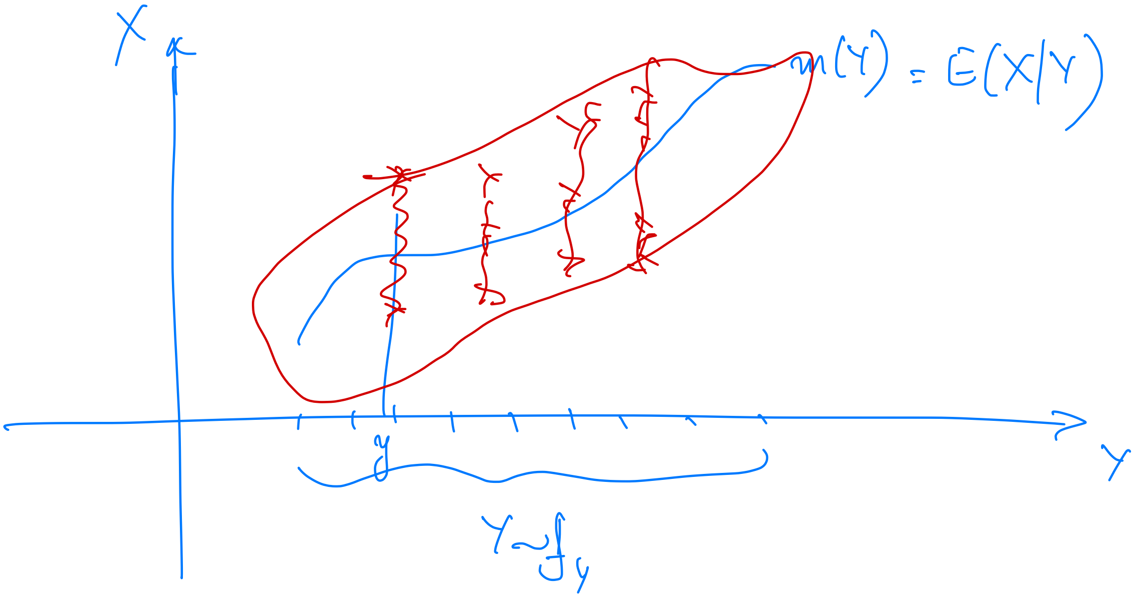
5-3-2023

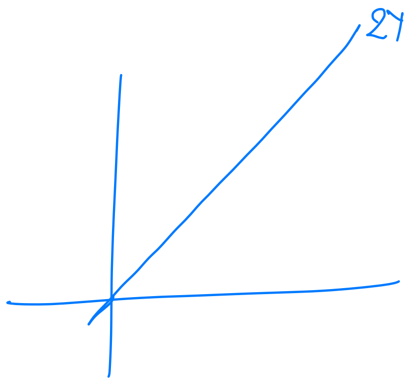
$$Var(X) = E\left(Var(X|Y)\right) + Var\left(\underbrace{E(X|Y)}_{\text{γνωστή συνάρτηση } m(Y)}\right)$$

$$E(X) = E\left(E(X|Y)\right) = E\left(m(Y)\right)$$

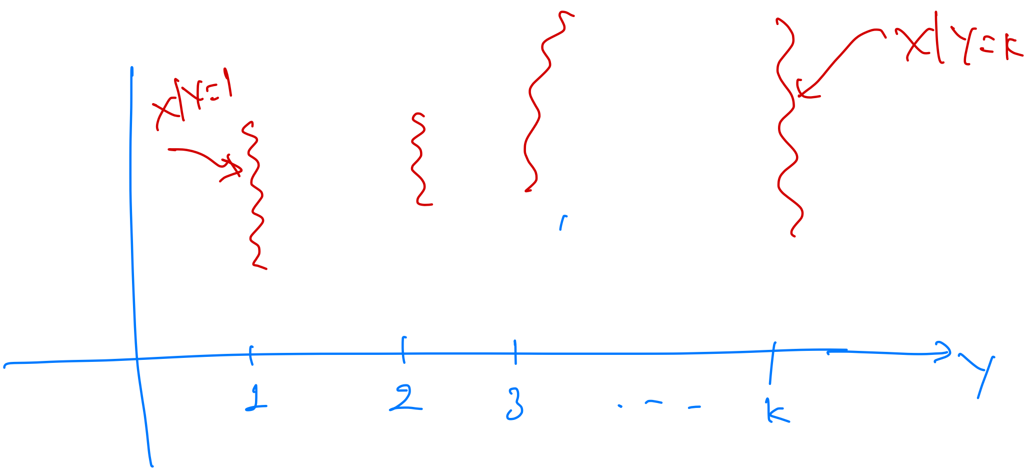


π.χ. αν $X = 2Y$

$E(X|Y) = 2Y$



$Y \in \{1, 2, \dots, k\}$ $\quad p_j = P(Y=j)$



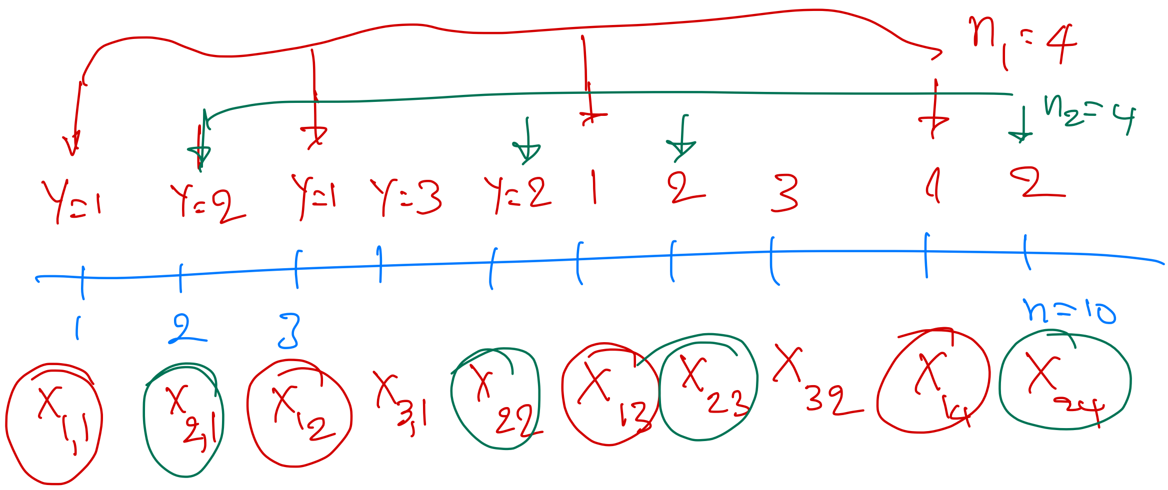
$$E(X) = \sum_{j=1}^{K} p_j \, E(X \mid Y=j)$$

Αρχική Monte Carlo:

1. Δημ. $Y \sim P$ Έστω $Y=j$
2. Αν $Y=j \to$ Δημ. $X \sim X \mid Y=j$

$$\bar{X}_n = \frac{1}{n}\sum_{t=1}^{n} X_t = \qquad \text{Έστω } n_j = \sum_{t=1}^{n} 1(Y_t=j)$$

$$= \frac{1}{n}\sum_{j=1}^{K}$$



$$\sum_{t=1}^{n} X_{j} = \sum_{j=1}^{K} \sum_{t=1}^{n_j} X_{j,t}$$

$$= \sum_{j=1}^{K} n_j \, \bar{X}_{j,n_j}$$

$$\Rightarrow \frac{1}{n}\sum_{t=1}^{n} X_t = \sum_{j=1}^{K} \underbrace{\frac{n_j}{n}}_{\downarrow\, p_j} \underbrace{\overline{X}_{j,n_j}}_{\downarrow\, E(X|Y=j)}$$

$$X_t = F_X$$

$$\hat{\theta} = \tilde{X} = \sum_{j=1}^{K} p_j \overline{X}_{j,n_j}$$

n_j "αυθαίρετα"

$n_j \longrightarrow \infty \quad \forall j$

Στρωματοποιημένη Δειγματοληψία (Stratified Sampling)

$n_j = ?$

---

① $Y = \begin{cases} 1 & p_1 \\ 2 & \vdots \\ \vdots & \\ k & p_k \end{cases}$

n παρατηρήσεις από Y

$E(\#\text{παρατ. } Y=j) = n p_j \longleftarrow n_j$

$\theta = EX$

$$\hat{\theta}_1 = \frac{1}{n} \sum_{t=1}^{n} X_t \qquad Var(\hat{\theta}_1) = \frac{1}{n} Var(X) \qquad (1)$$

Stratified Sampling

$$\hat{\theta}_2 = \sum_{j=1}^{K} p_j \overline{X}_{j,n_j}$$

$$Var(\hat{\theta}_2) \quad \sum_{j=1}^{K} p_j^2 \, Var(\overline{X}_{j,n_j})$$

$$= \sum_{j=1}^{K} p_j^2 \, \frac{\overbrace{Var(X|Y=j)}^{s_j^2}}{n_j} = \sum_{j=1}^{K} \frac{p_j^2}{n_j} s_j^2$$

(1) $n_j = n p_j \quad , j = 1, \dots, K$

$$\Rightarrow Var(\hat{\theta}_2) = \sum_{j=1}^{K} \frac{p_j^2}{n p_j} s_j^2 = \frac{1}{n} \sum_{j=1}^{K} p_j s_j^2$$

$$= \frac{1}{n} \cdot \underbrace{\sum_{j=1}^{K} P(Y=j) \, Var(X|Y=j)}$$

$$= \frac{1}{n} E_Y\big(Var(X|Y)\big) \qquad (2)$$

$$Var(\hat{\theta}_1) = \frac{1}{n} Var(X) \qquad (1)$$

$$Var(X) = E\big(Var(X|Y)\big) + Var\big(E(X|Y)\big)$$

αυτή τη διασπορά αποφεύγουμε

Γενικότερο ερώτημα $\quad n_j = ?$

έτσι ώστε $\quad \min \operatorname{Var} \hat{\theta}$

$$\sum_{j=1}^{k} n_j = n \quad ?$$

$$\operatorname{Var} \hat{\theta} = \sum_{j=1}^{k} \frac{p_j^2}{n_j} s_j^2$$

Π. μη γραμμ. πρ.

$$\left.\begin{aligned} &\min_{n_1,\dots,n_k} \sum_{j=1}^{k} \frac{p_j^2}{n_j} s_j^2 \\ &\sum_{j=1}^{k} n_j = n \\ &n_j \geq 0 \\ &\cancel{n_j \in \mathbb{Z}} \end{aligned}\right\}$$

$$L = \sum_{j=1}^{k} \frac{p_j^2}{n_j} S_j^2 - \mu \left(\sum_{j=1}^{k} n_j - n \right) \qquad \mu: \text{πολ. Lagrange}$$

$$\frac{\partial L}{\partial n_j} = -\frac{p_j^2 s_j^2}{n_j^2} - \mu = 0 \Rightarrow$$

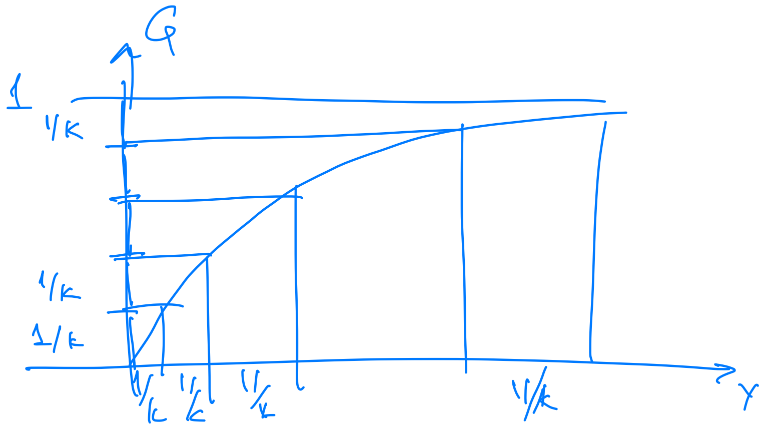
$$\Rightarrow n_j^2 = \left(-\frac{1}{\mu} \right) \cdot p_j^2 s_j^2 = c \, p_j^2 s_j^2 \Rightarrow$$

$$\Rightarrow n_j = \tilde{c} \cdot p_j S_j \qquad \forall j$$

$$\sum n_j = n \Rightarrow \tilde{c} = \frac{n}{\sum p_j s_j}$$

$$\Longrightarrow \quad n_j^* = n \cdot \frac{P_j S_j}{\sum_{j=1}^{n} P_j S_j}$$

Γενίκευση όταν $\quad Y \sim G \quad$ συνεχής κατανομή.



$Y: \; n_1, \; n_2, \dots, n_k$

Μείωση Διασποράς μέσω Importance Sampling

(Δειγματοληψία Σπουδαιότητας)

Έστω $X \sim f$ pdf

$$\theta = E_f(h(X)) = \int h(x) f(x)\,dx.$$

Έστω άλλη pdf $g(x)$

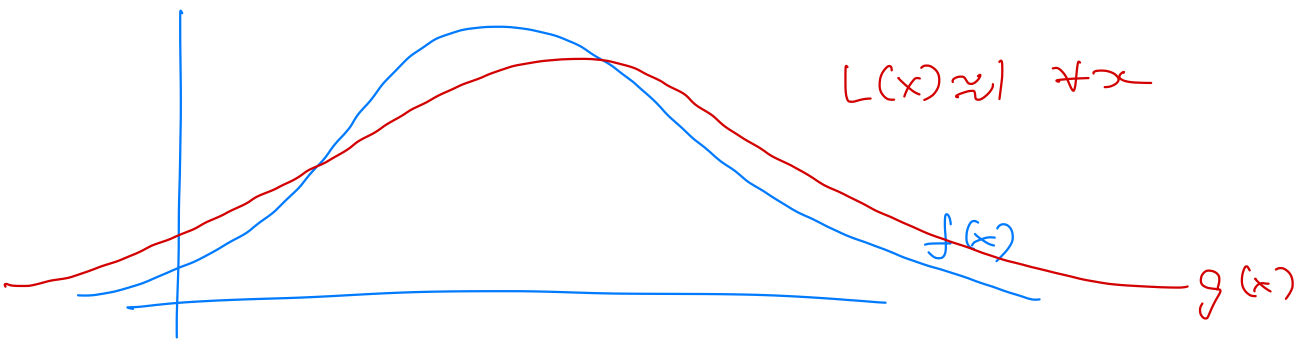
$$\theta = \int h(x) \cdot \frac{f(x)}{g(x)} \cdot g(x)\,dx = E_g\left[h(X) \cdot \frac{f(X)}{g(X)}\right]$$

$$= E_g\left[h(X)\,L(X)\right]$$

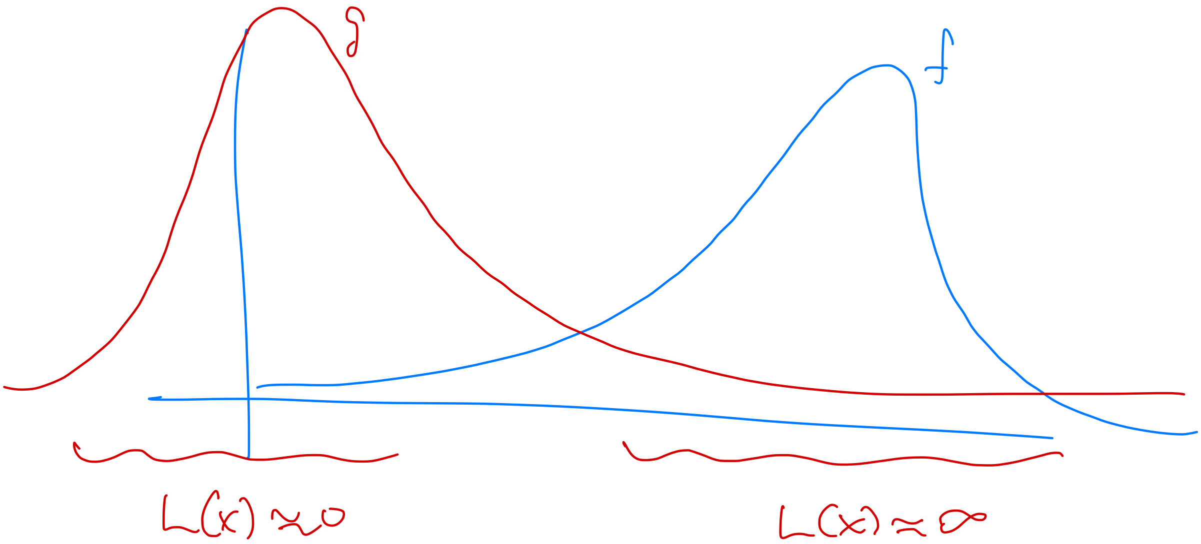
$$L(X) = \frac{f(X)}{g(X)}, \quad X \sim g$$

Πότε είναι χρήσιμο?

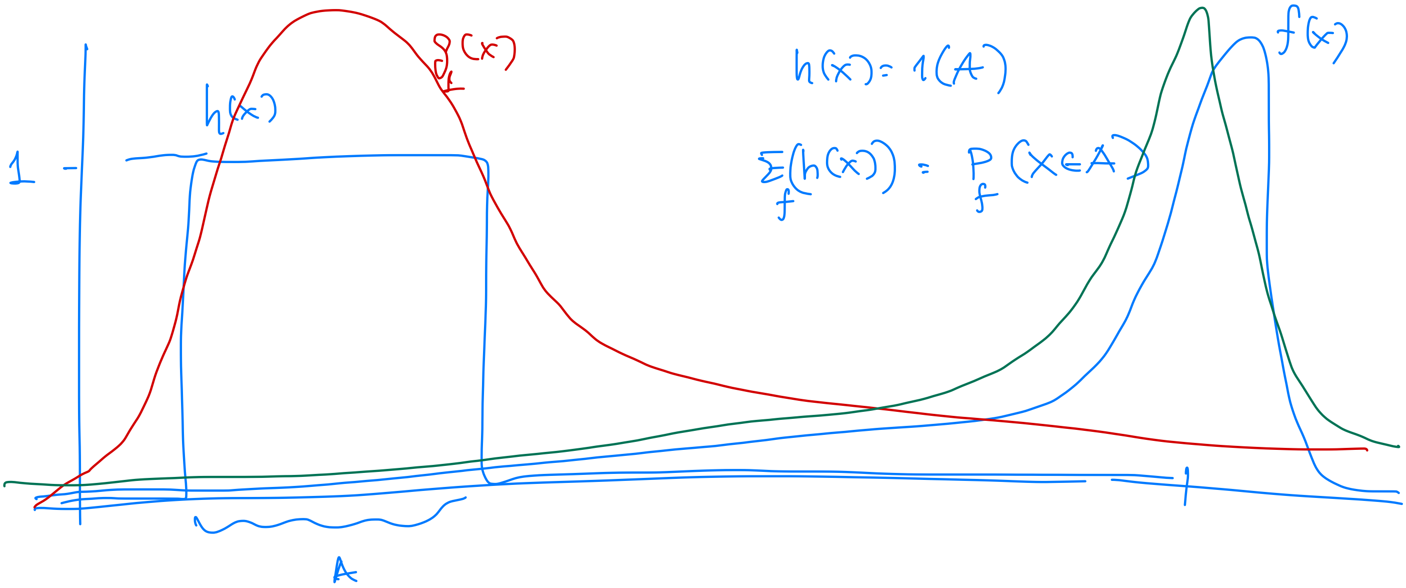
Παρατηρούμε $E_g(L(X)) = \int \frac{f(x)}{g(x)} g(x)\,dx = 1$



$Var(L(X))$ μικρή



$Var(L(X))$ πολύ μεγάλη

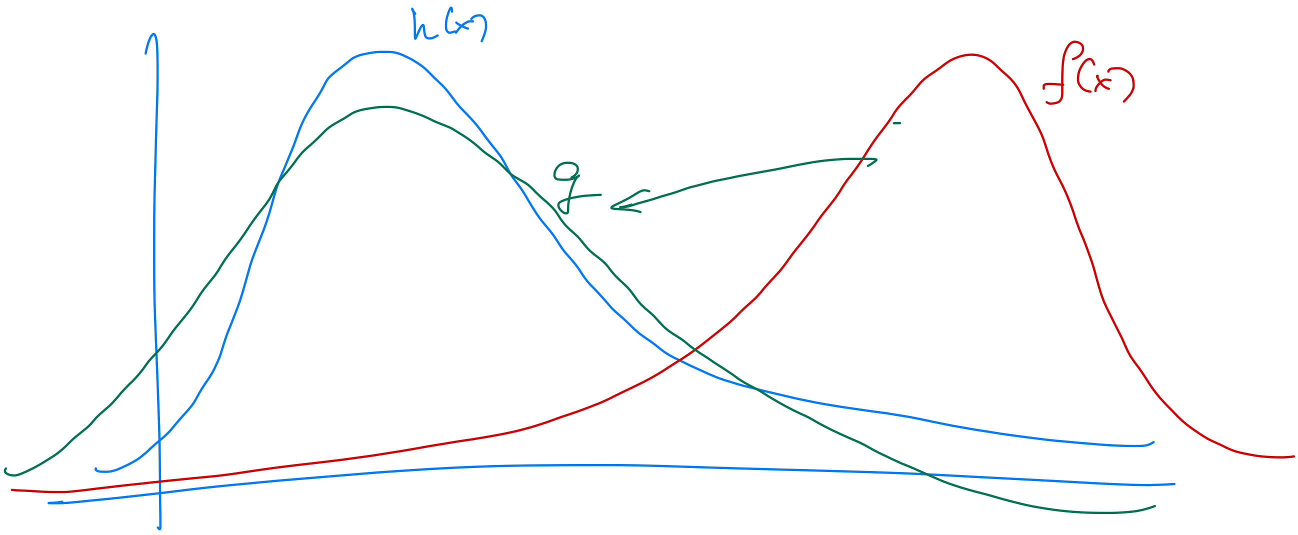


$h(x) = 1(A)$

$E_f(h(x)) = P_f(X \in A)$

$$\theta = E_{g_1}\left[\underbrace{h(X) \cdot \frac{f(x)}{g_1(x)}}_{\text{μικρές τιμές}} \right] \quad \forall x.$$

$$\theta = E_{g_2}\left[h(x) \cdot \frac{f(x)}{g_2(x)} \right]$$



h(x)
f(x)
g