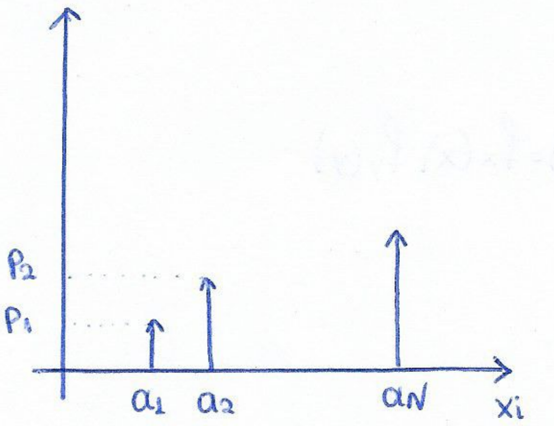
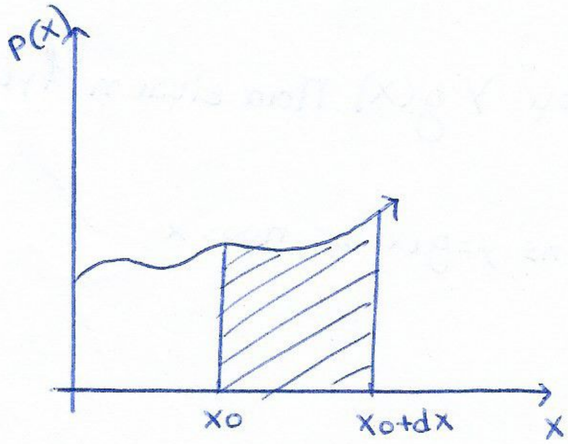
# Μάθημα 2

*** 1/3/18 ***

Διακριτή τυχαία μεταβλητή A




Συνεχής τ.μ. X




$$P(x\in(a,b))=\int_a^b p(x)dx$$

$p(x)$: Συνάρτηση πυκνότητας πιθανότητας (pdf)

$$0\le p(x)\le 1$$

$$\int_{-\infty}^{+\infty} p(x)dx=1$$

Ορίζω για τ.μ να πάρει τιμή μέσα σ' ένα διάστημα

Μέση τιμή τ.μ. X: $E(x)=\int_{-\infty}^{+\infty} x f_x(x)dx$

Μεταβλητότητα τ.μ. x

$$Var[x]=E[x^2]-(E[x])^2=E[(x-E(x))^2]=\sigma^2$$

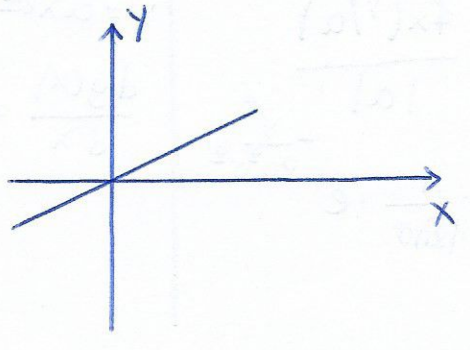
Συμμεταβλητότητα τ.μ. X

$$cov(x,y)=E[(x-E(x))(y-E(y))]$$

Συντελεστής συσχέτισης

$$\rho(x,y)=\frac{cov(x,y)}{\sigma_x\cdot\sigma_y} \quad -1\le\rho(x,y)\le 1$$

Για τον συντελεστή συσχέτισης



## Πολυδιάστατες τ.μ. (συνεχείς)

τ.μ. X, Y  $f_{xy}(x,y) \rightarrow$ από κοινού pdf

$f_{x/y}(x) \rightarrow$ δεσμευμένη pdf

Αν X, Y ανεξάρτητες $\rightarrow f_{xy}(x,y) = f_x(x) \cdot f_y(y)$

Κανόνας Bayes

$$f_{xy}(x,y) = f_{y/x}(y) \cdot f_x(x)$$

$$f_x(x) = \int_{-\infty}^{+\infty} f_{xy}(x,y)\,dy$$

## Μετασχηματισμοί τ.μ

τ.μ. X με $f_x(x)$ μετασχηματίζεται στην τμ $Y: g(X)$. Ποια είναι η $f_Y(y)$;

$$f_Y(y) = \sum_{i=1}^{N} \frac{f_x(x_i)}{\left|\frac{dg(x)}{dx}\right|_x}$$ , με $x_i$ οι λύσεις της $y = g(x)$ ως προς x

### Παράδειγμα

$X \sim N(0,\sigma) \rightarrow$ κανονική κατανομή με μέση τιμή 0 και τυπική απόκλιση σ

$Y = a \cdot X$  $f_Y(y) = ?$

Γενικά $X \sim N(\mu,\sigma)$ τότε

$$f_x(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

Εδώ $f_x(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$

$$f_Y(y) = \frac{f_x(y/a)}{|a|}$$

$$= \frac{1}{|a|} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{y^2}{2a^2\sigma^2}}$$

$y = ax \Leftrightarrow x = \frac{y}{a}$

$\frac{dg(x)}{dx} = \frac{dy}{dx} = a$

## Παράδειγμα

τ.μ. $X \sim N(0,1)$

τ.μ. $Y = a \cdot X^2$

$f_Y(y) =;$

### Λύση

$$y = ax^2 \Rightarrow x = \pm\sqrt{\frac{y}{a}} \quad , \quad \frac{dy}{dx} = 2ax$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$f_Y(y) = \sum_{i=1}^{2} \frac{f_X(x)}{\left|\frac{dy}{dx}\right|} = \frac{f_X\left(\sqrt{\frac{y}{a}}\right)}{\left|2a\sqrt{\frac{y}{a}}\right|} + \frac{f_X\left(-\sqrt{y/a}\right)}{\left|-2a\sqrt{y/a}\right|}$$ , Θεωρούμε $a>0$

$$= \frac{1}{2\sqrt{2\pi a y}}\left[e^{-\frac{y}{2a}} + e^{-\frac{y}{2a}}\right] = \frac{1}{\sqrt{2\pi a y}} e^{-\frac{y}{2a}}$$

## Μετασχηματισμός πολυδιάστατων τ.μ.

τ.μ. X και Y με $f_{XY}(x,y)$

Μετασχηματισμός

$$z = g(x,y)$$
$$w = h(x,y)$$

$$f_{ZW}(z,w) = \sum_{i=1}^{N} \frac{f_{XY}(x,y)}{|J(x,y)|_{(x,y)}}$$

όπου $x_i, y_i$ λύσεις του συστήματος $\begin{cases} z = g(x,y) \\ w = h(x,y) \end{cases}$

και $\underbrace{J(x,y)}_{\text{Jacobi}} = \begin{vmatrix} dg/dx & dg/dy \\ dh/dx & dh/dy \end{vmatrix}$

## Παράδειγμα

τ.μ. X, Y με $f_{XY}(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$

$Z = X + Y$

$W = X - Y$

$f_{ZW}(z,w) = ?$

### Λύση

Λύση του συστήματος $\begin{cases} z = x+y \\ w = x-y \end{cases} \Leftrightarrow \begin{cases} x = \frac{z+w}{2} \\ y = \frac{z-w}{2} \end{cases}$

$$J = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2 \Rightarrow |J| = 2$$

Άρα $f_{ZW}(z,w) = \dfrac{f_{XY}\left(\frac{z+w}{2}, \frac{z-w}{2}\right)}{2} = \cdots = \frac{1}{4\pi} e^{-\frac{z^2+w^2}{4}}$

## Κατανομές διακριτών τ.μ

- Bernoulli τ.μ. $R \in \{0,1\}$ π.χ. ρίψη κέρματος
  $p(R=1) = f$, $p(R=0) = 1-f$ $\quad 0 \le f \le 1$
  $Bern(r; f) = f^r (1-f)^{1-r}$
  $E[r] = f$
  $Var[r] = f(1-f)$
- Διωνυμική κατανομή (N πραγματοποιήσεις της Bernoulli)
  $Bin(r; f, N) = \binom{N}{r} f^r (1-f)^{N-r}$
  $E[r] = N \cdot f$
  $Var[r] = N \cdot f (1-f)$

$\sum_{k=0}^{n} \binom{n}{k} = 2^n$

Χωρίς διάταξη

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Με διάταξη

$\binom{n}{k} = \frac{n!}{(n-k)!}$

$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}$

## Συνεχείς τ.μ.

### • Πολυδιάστατη κανονική κατανομή

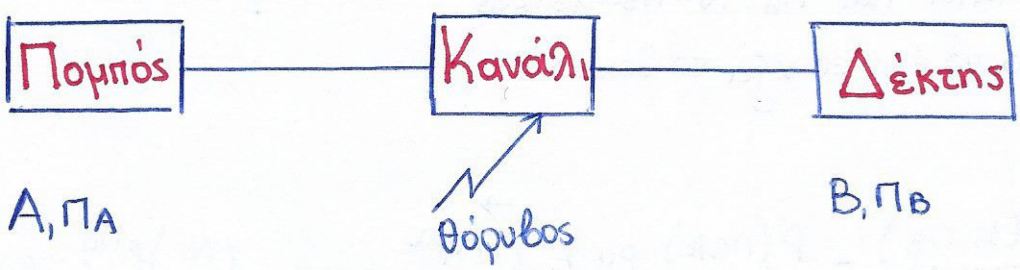
τ.μ. $X=[x_1 \ x_2 \ \ldots \ x_d]^T$

$$p(x)=\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}$$

Σ: πίνακας
μ: μέση τιμή τυχαίων συνιστωσών

$\mu = E(x)$

Σ: πίνακας μεταβλητότητας: $E[(x-\mu)(x-\mu)^T]$




## Ευθύ και αντίστροφο πρόβλημα

### Ευθύ πρόβλημα

- Μοντέλο παραγωγής παρατηρήσεων
- Στόχος: εκτίμηση της pdf και των παραμέτρων της τ.μ.

### Παράδειγμα

Δοχείο με K μπάλες
B μαύρες
W=K−B λευκές
Επιλέγουμε N φορές μια μπάλα και επανατοποθετούμε.

α) Ποια η pdf του $n_B$="Αριθμός των φορών που επιλέχτηκε μαύρη μπάλα"

β) $E[n_B]=;$ , $\sigma_{n_B}=Var[n_B]=;$

#### Λύση

Ορίζω $f_B=\frac{B}{K}$ (η πιθανότητα να τραβήξω μαύρη μπάλα)

α) Διωνυμική $p(n_B, f, N)=\binom{N}{n_B}(1-f_B)^{N-n_B}$

β) $E[n_B]=N \cdot f_B$ , $\sigma_{n_B}=N f_B(1-f_B)$

## Αντίστροφο πρόβλημα

- Μοντέλο παραγωγής παρατήρησης
- Στόχος: εκτίμηση pdf "κρυμμένων" μεταβλητών
- Μπεϋζιανή μεθοδολογία → Χρήση του κανόνα του Bayes

### Παράδειγμα

11 δοχεία $k \in \{0, 1, \dots, 10\}$

Κάθε δοχείο k περιέχει 10 μπάλες.

Το δοχείο k περιέχει k μαύρες και 10-k λευκές

Επιλέγουμε ένα δοχείο (δεν γνωρίζω ποιο)

Επιλέγουμε N φορές μια μπάλα και επανατοποθετούμε.

Έχω $n_B$ μαύρες μπάλες και $n_H = N - n_B$ λευκές.

Ποια η πιθανότητα να έχω επιλέξει το δοχείο k;

Έστω ότι είναι ομοιόμορφη κατανομή

**Λύση**

$$p(k|n_B; N) = \frac{P(k, n_B)}{P(n_B)} = \frac{P(n_B|k) \cdot p(k)}{P(n_B)}$$

$p(k)$: εκ των προτέρων πιθανότητα

$p(k|n_B)$: εκ των υστέρων πιθανότητα

Έστω $p(n_B|k) = \binom{N}{n_B} f_k^{n_B} (1-f_k)^{N-n_B}$

$f_k = \frac{k}{10}$

$P(k) = \frac{1}{11}$ (ομοιόμορφα)

$$p(k|n_B) = \frac{P(n_B|k) P(k)}{P(n_B)}$$

$P(n_B|k) \to$ διωνυμική

$P(k) = 1/11$ ομοιόμορφη

$$p(n_B) = \sum_k p(n_B, k) = \sum_k p(k) \cdot p(n_B|k)$$

Άρα

$$p(k|n_B) = \frac{1}{11} \frac{\binom{N}{n_B} f_k^{n_B} (1-f_k)^{N-n_B}}{\sum_k \binom{N}{n_B} f_k^{n_B} (1-f_k)^{N-n_B}}$$

Αριθμητικό παράδειγμα $(N=10, n_B=3)$

| k | $p(k\|n_B;N)$ | |
|---|---|---|
| 0 | 0 | $\to f_k = 0$ |
| 1 | 0,063 | |
| 2 | 0,22 | |
| 3 | 0,29 | |
| 4 | 0,24 | |
| 5 | 0,13 | |
| 6 | 0,047 | |
| 7 | 0,0099 | |
| 8 | 0,00086 | |
| 9 | 0,0000096 | |
| 10 | 0 | $\to f_k = 1$ |

Έχω $n_B=3, N=10$

Ποια η πιθανότητα στην επόμενη επιλογή $(11^{η})$ η μπάλα να είναι μαύρη.

$$p(ball_{N+1}=black|n_B,N)=\sum_k p(ball_{N+1}=black,k|n_B,N)$$

$$=\sum_k \underbrace{p(ball_{N+1}=black|k,n_B,N)}_{\text{ανεξάρτητο του } n_B} \cdot p(k|n_B,N)$$

$$=\sum_k \frac{k}{10}p(k|n_B;N) \Rightarrow$$

$$p(ball_{N+1}=black|n_B,N)=\sum_k \frac{k}{10}p(k|n_B;N)$$

$N=10,\ n_B=3$

$$p(ball_{N+1}=black|n_B=3,N=10)=0{,}3333$$

Χρησιμοποιήσαμε Μπεϋζιανή θεώρηση άρα περιθωριοποίηση του k (της κρυφής μεταβλητής)

Αν ~~κ~~ κάναμε στατιστική θεώρηση
↳ Επιλέγω από το $p(k|n_B,N)$ τη μέγιστη
Εδώ $k=3 \to p(n_B|k)=\frac{3}{10}=0{,}3$

→ Η στατιστική θεώρηση δίνει τιμή στην κρυφή μεταβλητή

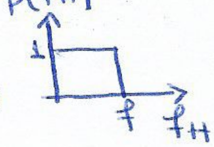
Παράδειγμα (Συνεχής "κρυμμένη" μεταβλητή)

Ρίχνουμε κέρμα N φορές ("πειραγμένο" κέρμα) με $f_H=p(\text{"Heads"})\neq 1/2$

- παρατηρούμε $n_H$ φορές το H
- ποια η πιθανότητα του $f_H$ $(0\leq f_H\leq 1)$ $\longrightarrow p(f_H/n_H)$

παράμετρος

- Μετατρέπουμε το $f_H$ σε τμ $\to p(f_H)=1$ ομοιόμορφη




Λύση $$p(f_H/n_H;N)=\frac{p(n_H/f_H)\cancelto{1}{p(f_H)}}{p(n_H)}=\frac{p(n_H/f_H)}{p(n_H)}=\frac{p(n_H/f_H)}{\int p(n_H,f_H)df_H}=$$

$$\frac{p(n_H/f_H)}{\int p(n_H/f_H)p(f_H)df_H}$$

$$p(n_H|f_H) = \binom{N}{n_H} f^{n_H}(1-f)^{N-n_H}$$

→ ολοκλήρωση Beta

$$p(n_H) = \int_0^1 p(n_H/f_H)\,df_H = \int_0^1 \binom{N}{n_H} f^{n_H}(1-f)^{N-n_H}\,df_H =$$

$$\binom{N}{n_H}\int_0^1 f^{n_H}(1-f)^{N-n_H}\,df_H = \binom{N}{n_H}\frac{n_H!(N-n_H)!}{N+1} = \frac{1}{N+1}$$

$$\Rightarrow P(f_H/n_H, N) = (N+1)\binom{N}{n_H} f^{n_H}(1-f)^{N-n_H}$$

Ερώτημα (συνέχεια από το προηγούμενο)

→ κανόνας της άθροισης

$$p(\text{ρίψη}_{N+1}=\text{"H"}\,|\,n_H, N) = \int_0^1 P(\text{Ρίψη}=\text{"H"}, f_H\,|\,n_H, N)\,df_H$$

$$= \int_0^1 \underbrace{P(\text{Ρίψη}=\text{"H"}/f_H, n_H, N)}_{\text{ανεξάρτητο του } n_H} \cdot P(f_H/n_H, N)\,df_H$$

$$= \int_0^1 f_H\, P(f_H/n_H, N)\,df_H = \dots = \frac{n_H+1}{N+1}$$

Χρησιμοποιώ το τυπολόγιο της διωνυμικής και την ολοκλήρωση Beta