042_Folien

November 23, 2018

```
In [1]: import numpy as np                      # mathematical methods
        from scipy import stats                 # statistical methods
        from matplotlib import pyplot as plt    # plotting methods
        %matplotlib inline
```

0.1 Diskrete Wahrscheinlichkeitstheorie - Überblick

1. Zufall

- Zufallsvariable, Zufallsexperiment, Wahrscheinlichkeitsmaß, Wahrscheinlichkeitsverteilung, Axiome von Kolmogorov

2. Zusammengesetzte Wahrscheinlichkeiten

- bedingte Wahrscheinlichkeit, gemeinsame Wahrscheinlichkeit, totale Wahrscheinlichkeit, Satz von Bayes

3. Kenngrößen

- Erwartungswert & Varianz

4. Modellverteilungen

- eindimensional

multivariate Verteilungen

multivariate Verteilungen

- bivariate Verteilung
- zusammengesetzte Verteilung
- Erwartungswert
- Varianz
- Kovarianz
- stochastische Unabhängigkeit

1 Zweidimensionale Verteilungen: *bivariat*

- vektorielle Zufallsvariable $\boldsymbol{W} = (X, Y)$

Beispiele

- roter und grüner Würfel gleichzeitig geworfen
- Roulette *rot* und *ungerade*
- Koordinaten (x_i, y_i)
- Alter und Blutdruck bei Patienten
- Pixel Helligkeit und Farbe

1.1 Ein Ergebnis (Elementarereignis)

vektoriell

$$\boldsymbol{w} = (x, y)$$

1.2 Verbundene Wahrscheinlichkeit

$$P(w) \;=\; P(X = x \cap Y = y) \;=\; p(x, y) \;=\; p_{xy}$$

1.3 Stochastische Unabhängigkeit

$$P(X = x \cap Y = y) \;=\; P(X = x) \cdot P(Y = y)$$

1.4 Erwartungswert

$$\boldsymbol{\mu} = \begin{pmatrix} \mathcal{E}(X) \\ \mathcal{E}(Y) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N_x} x_i \cdot P(X = x_i) \\ \sum_{j=1}^{N_y} y_j \cdot P(Y = y_j) \end{pmatrix}$$

1.4.1 mit den Randverteilungen

$$P(X = x_i) = \sum_{j=1}^{N_y} P(X = x_i \cap Y = y_j)$$

$$P(Y = y_i) = \sum_{i=1}^{N_x} P(X = x_i \cap Y = y_j)$$

1.4.2 Kontingenztafel

Am Beispiel zweier Münzen

Wahrscheinlichkeit p_{xy} und Randverteilungen p_x und p_y

```
M1\M2 |  0      1      |
----------------------------
0     |  1/4    1/4    | 1/2
1     |  1/4    1/4    | 1/2
----------------------------
      |  1/2    1/2    |  1
```

***Ein* mögliches Beispiel für Werte x_{ij}**

```
M1\M2 |  0      1      |
-----------------------
0     |  0      1      |
1     |  1      4      |
```

1.5 Erwartungswert einer Summe von Zufallsvariablen

Für zwei Zufallsvariablen X und Y ist $Z = X + Y$ wieder eine Zufallsvariable. Es gilt

$$\mathcal{E}(Z) = \mathcal{E}(X + Y) = \mathcal{E}(X) + \mathcal{E}(Y)$$

Beweis

$$\mathcal{E}(X + Y) = \sum_{i=1}^{N} (x_i + y_i) p_i = \sum_{i=1}^{N} x_i p_i + \sum_{i=1}^{N} y_i p_i = \mathcal{E}(X) + \mathcal{E}(Y)$$

Was sind i und N?

Beispiel 2 Münzen, Anzahl "*Kopf*"; erste: $i \quad X \quad p\, 1 \quad 0 \quad 1/2$ $\qquad \mathcal{E}(X) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}\, 2 \quad 1 \quad 1/2$

zweite: Dasselbe für Y: $\quad \mathcal{E}(Y) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$

Nun die Summe (Anzahl "Zahl") der zwei Münzen i Z p 1 0 1/4 2 1 1/2 3 2 1/4 $\mathcal{E}(Z) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$

Gemeinsames Werfen von Münze X und Münze Y ergibt Summe "Anzahl *Kopf*" Z.
Elementarzerlegung des gemeinsamen Vektors $\boldsymbol{W}$: $\forall i \in \{1 \dots N\}$ gilt $z_i = x_i + y_i$.

```
i    X    Y    Z=X+Y   p
1    0    0    0       1/4
2    0    1    1       1/4
3    1    0    1       1/4
4    1    1    2       1/4
```

4 Kombinationsmöglichkeiten, N=4, alle durchlaufen mit i

$$\begin{aligned}
\mathcal{E}(X+Y) &= 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} &&= 1 \\
\mathcal{E}(X) &= 0 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} &&= \frac{1}{2} \\
\mathcal{E}(Y) &= 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} &&= \frac{1}{2} \\
\mathcal{E}(X) + \mathcal{E}(Y) &= 2 \cdot \frac{1}{2} &&= 1
\end{aligned}$$

```
X \ Y |   0       1      |
----------------------------
0     |  1/4     1/4     | 1/2
1     |  1/4     1/4     | 1/2
----------------------------
      |  1/2     1/2     |  1
```

$$\sum_{k=1}^{4} (x_k + y_k) \cdot p_k = \sum_{i=1}^{2} \sum_{j=1}^{2} (x_i + y_j) p_{ij}$$

Damit:

$$= \sum_{i=1}^{2} x_i \sum_{j=1}^{2} p_{ij} + \sum_{j=1}^{2} y_j \sum_{i=1}^{2} p_{ij}$$

$$= \sum_{i=1}^{2} x_i p_i + \sum_{j=1}^{2} y_j p_j$$

mit $k = 2 \cdot (i-1) + j$

1.6 Anwendung bi- / multivariate Verteilung:

Beispiele

- Verkehrszählung nach Verkehrsmittel und Gesamtaufkommen
- Schadstoffausstoß und Gesamtbelastung
- Vogelzug mehrerer Arten pro Tag und Gesamtanzahl

1.6.1 Addition Poisson-Verteilungen

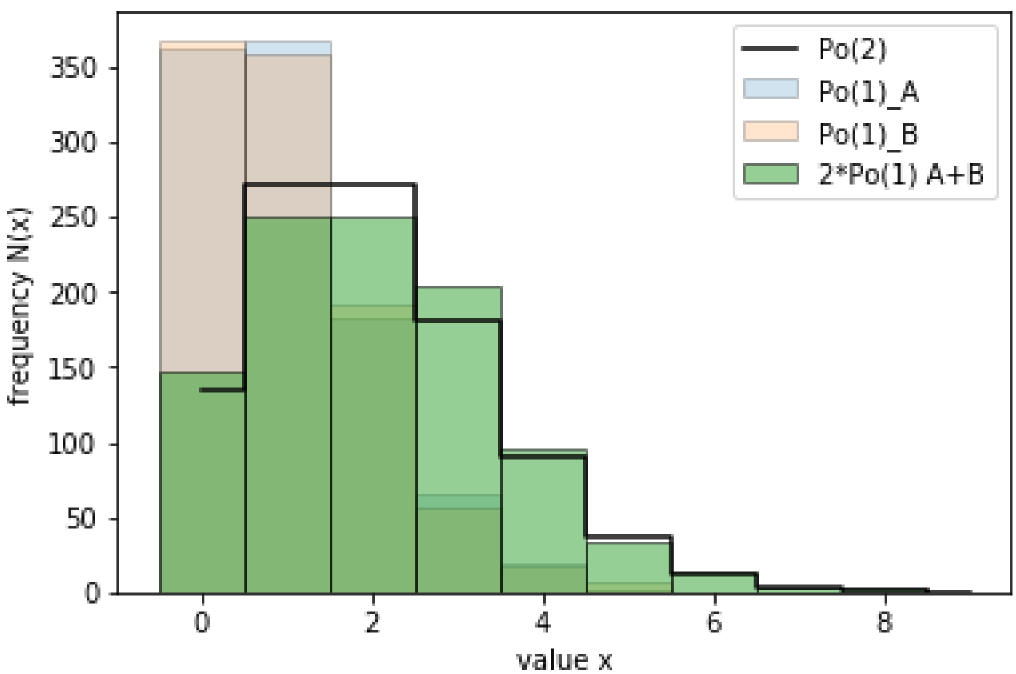
Sind $X \sim \mathcal{P}\wr(\lambda_1)$ und $Y \sim \mathcal{P}\wr(\lambda_2)$ Poisson-verteilt und voneinander unabhängig, so ist

$$X + Y \sim \mathcal{P}\wr(\lambda_1 + \lambda_2)$$

In [3]:
```
'''test summation of Poisson distributions
   according to rate lambda
'''
l1 = 1.                              # lambda of distribution A and B
x = np.arange(0., 21)                # outcomes
N = 1000                             # number of samples to draw
np.random.seed(987654)
xa = stats.poisson(l1).rvs(size=N)   # outcomes "A" of Poisson(lambda=1)
xb = stats.poisson(l1).rvs(size=N)   # outcomes "B" of Poisson(lambda=1)
```

```
    xab = xa + xb                       # sum of the two
    print('Po(1)_A = ', xa[:30], '...')
    print('Po(1)_B = ', xb[:30], '...')
    print('sum A+B = ', xab[:30], '...')
    # compare the two single distributions, theire sum and the theoretical
    ha  = plt.hist(xa,  bins=np.arange(10), alpha=0.2, label='Po(1)_A',
                   align='left', edgecolor='black')
    hb  = plt.hist(xb,  bins=np.arange(10), alpha=0.2, label='Po(1)_B',
                   align='left', edgecolor='black')
    hab = plt.hist(xab, bins=np.arange(10), alpha=0.5, label='2*Po(1) A+B',
                   align='left', edgecolor='black')
    po2 = stats.poisson.pmf(x, 2*l1) # theoretical distribution with double lambda
    plt.step(x[:10], N*po2[:10], 'k--', where='mid', label='Po(2)')
    plt.xlabel('value x')
    plt.ylabel('frequency N(x)')
    plt.legend();
```

```
Po(1)_A =  [0 1 0 1 0 1 2 0 0 1 2 0 0 3 2 0 1 0 0 2 2 0 0 0 0 2 2 0 1 1] ...
Po(1)_B =  [1 0 0 1 1 1 0 0 2 1 1 0 0 2 2 0 0 0 0 3 1 0 0 1 0 1 2 0 2 3] ...
sum A+B =  [1 1 0 2 1 2 2 0 2 2 3 0 0 5 4 0 1 0 0 5 3 0 0 1 0 3 4 0 3 4] ...
```




```
In [6]: '''compare theoretical Poisson(2) with sum of two Poisson(1) distributions'''
        po11 = np.asarray([
                np.asarray([
                    y*x for j,y in enumerate(po1) for k,x in enumerate(po1) if j+k==i]).sum()
                for i in range(10)])
        for i in range(10):
            print('[P1+P1](x={})={:9.7f}  P2(x={})={:9.7f}'.format(i, po11[i], i, po2[i]))
```

```
[P1+P1](x=0)=0.1353353  P2(x=0)=0.1353353
[P1+P1](x=1)=0.2706706  P2(x=1)=0.2706706
[P1+P1](x=2)=0.2706706  P2(x=2)=0.2706706
[P1+P1](x=3)=0.1804470  P2(x=3)=0.1804470
[P1+P1](x=4)=0.0902235  P2(x=4)=0.0902235
[P1+P1](x=5)=0.0360894  P2(x=5)=0.0360894
```

```
[P1+P1](x=6)=0.0120298  P2(x=6)=0.0120298
[P1+P1](x=7)=0.0034371  P2(x=7)=0.0034371
[P1+P1](x=8)=0.0008593  P2(x=8)=0.0008593
[P1+P1](x=9)=0.0001909  P2(x=9)=0.0001909
```

1.7 Erwartungswert eines Produkts von Zufallsvariablen

Für zwei **unabhängige** Zufallsvariablen X und Y gilt:

$$\mathcal{E}(X \cdot Y) = \mathcal{E}(X) \cdot \mathcal{E}(Y)$$

Beweis ÜA

Beispiel zwei Münzen

```
i  j  k   X   Y   X*Y  p=px*py
1  1  1   0   0   0    1/4
1  2  2   0   1   0    1/4
2  1  3   1   0   0    1/4
2  2  4   1   1   1    1/4
```

$$\begin{aligned}
\mathcal{E}(X) &= 0 \cdot \frac{1}{4} \cdot 2 + 1 \cdot \frac{1}{4} \cdot 2 &&= \frac{1}{2} \\
\mathcal{E}(Y) &= 0 \cdot \frac{1}{4} \cdot 2 + 1 \cdot \frac{1}{4} \cdot 2 &&= \frac{1}{2} \\
\mathcal{E}(X \cdot Y) &= 0 \cdot \frac{1}{4} \cdot 3 + 1 \cdot \frac{1}{4} &&= \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2}
\end{aligned}$$

2 Varianz(en) mehrdimensionaler Verteilung

$$\mathrm{Var}(\boldsymbol{W}) = \begin{pmatrix} \mathrm{Var}(X) \\ \mathrm{Var}(Y) \end{pmatrix} = \begin{pmatrix} \sum_x (x - \mathcal{E}(X))^2 p_x) \\ \sum_y (y - \mathcal{E}(Y))^2 p_y) \end{pmatrix}$$

3 Kovarianz

Definition:

$$\mathrm{Cov}(X,Y) := \mathcal{E}\Big((X - \mathcal{E}(X))(Y - \mathcal{E}(Y)) \Big) = \sum_x \sum_y (x - \mu_x)(y - \mu_y) \cdot p_{xy}$$

Verschiebungssatz

$$\mathrm{Cov}(X,Y) = \mathcal{E}(X \cdot Y) - \mathcal{E}(X) \cdot \mathcal{E}(Y)$$

Varianz

$$\mathrm{Var}(X) = \mathrm{Cov}(X,X)$$

Beweis: **[ÜA]**

3.0.1 Bedeutung der Kovarianz für die Varianz einer Summe von Zufallsvariablen:

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathcal{E}\big((X + Y - \mathcal{E}(X) - \mathcal{E}(Y))^2\big) \\
&= \mathcal{E}\big((X - \mathcal{E}(X))^2 + (Y - \mathcal{E}(Y))^2 + 2(X - \mathcal{E}(X)) \cdot (Y - \mathcal{E}(Y))\big) \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)
\end{aligned}$$

3.0.2 Extremfälle der Kovarianz für die Varianz einer Summe von Zufallsvariablen

vollständige Korrelation X, Y=X

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y) \\
&= \mathrm{Var}(X)+\mathrm{Var}(Y)+2(\mathcal{E}(XY)-2\mathcal{E}(X)\mathcal{E}(Y)) \\
&= \mathrm{Var}(X)+\mathrm{Var}(X)+2\big(\mathcal{E}(X^2)-\mathcal{E}^2(X)\big) \\
&= 4\mathrm{Var}(X) = \mathrm{Var}(2X)
\end{aligned}$$

vollständige Antikorrelation X, Y=-X

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathrm{Var}(X)+\mathrm{Var}(Y)+2\mathrm{Cov}(X,Y) \\
&= \mathrm{Var}(X)+\mathrm{Var}(Y)+2\big(\mathcal{E}(XY)-\mathcal{E}(X)\mathcal{E}(Y)\big) \\
&= \mathrm{Var}(X)+\mathrm{Var}(X)-2\big(\mathcal{E}(X^2)-\mathcal{E}^2(X)\big) \\
&= (2-2)\mathrm{Var}(X) = \mathrm{Var}(0X)
\end{aligned}$$

4 Korrelationskoeffizient

Definition:

$$\mathrm{Korr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(X,Y)}{\sigma_x\sigma_y}$$

Eigenschaften

- Korr $\in [-1,1]$
- Korr $=+1$ wenn $X=Y$
- Korr $=-1$ wenn $X=-Y$
- Korr $=0$ wenn X,Y stochastisch unabhängig
  - (nicht unbedingt umgekehrt)

5 Stochastische Unabhängigkeit

X und Y sind genau dann stochastisch unabhängig, wenn

$$\forall(x,y)\in\Omega: \quad P(X=x\cap Y=y) = P(X=x)\cdot P(Y=y)$$

$\Rightarrow$

$$\mathrm{Cov}(X,Y)=0$$

bzw.

$$\mathcal{E}(X\cdot Y) = \mathcal{E}(X)\cdot\mathcal{E}(Y)$$

6 Zusammenfassung multivariate Wahrscheinlichkeitsverteilung

- Eine Wahrscheinlichkeit auf mehrere Zufallsvariable, z.B. $p(x,y)$
- Erwartungswerte
- Varianzen
- Kovarianz
- Korrelation
- Unabhängigkeit

6.1 Mehr als zwei Dimensionen?

- Erwartungswert $\mathcal{E}(\boldsymbol{X})$
- Varianz $\mathrm{Var}(\boldsymbol{X})$
- Matrix gegenseitiger Abhängigkeiten
  - Wie bei empirischer Statistik "2D-Scatterplots"
- Kovarianzmatrix
  - Diagonale: Varianzen
  - Seiten: Kovarianzen

7 Ausblick

Kontinuierliche Verteilungen

8 Fragen?