確率・統計（電子２年）　　第８講

- 確率変数の期待値と分散（＝**分布の期待値と分散**）
- 中間レポート（前半模擬テスト）課題と解答用紙→第 12 講で解説・回収

前回復習

C 言語の標準ライブラリの関数 $drand48()$（$[0.0, 1.0)$ の範囲の $double$ 型乱数を返す）と $log1p(x)$（$double$ 値 x を貰って，$double$ 値 $\log(1+x)$ を返す．ただし，e を底）を使って，パラメタ $\lambda = 0.5$ の指数分布に従う $double$ 型乱数を生成するプログラムを作成して実験せよ．

前回学んだように，ある分布の分布関数を F（ただし逆関数を持つ）として，

- $X(\omega)$ が $[0, 1]$ 上の一様分布に従う確率変数の実現値（乱数）ならば，
- $F^{-1}(X(\omega))$ は分布 F に従う確率変数の実現値（乱数）になる．

パラメタ λ の指数分布の分布関数は，$F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$ なので，その逆関数は $F^{-1}(x) = -\frac{1}{\lambda}\log(1-x)$．よって，

- 確率変数 X が $[0, 1]$ 上の一様分布に従う時，$Y = F^{-1}(X) = -\frac{1}{\lambda}\log(1-X)$ は，パラメタ λ の指数分布に従う確率変数になる．

例えば，$\lambda = 0.5$ の指数分布に従う 1000 個の乱数（$Y_1(\omega), \ldots, Y_{1000}(\omega)$）を生成する C 言語のプログラム例は以下．なお，`srand48(99)` は乱数の種の初期化．

```
#include <math.h>
#include <stdlib.h>
#include <stdio.h>
main() {
  int i;  srand48(99);
  for (i=0;i<1000;i++) { printf("%f\n", -2*log1p(-drand48())); }
}
```

11. 確率変数の期待値と分散（参考書１．６－１．７）

確率変数の期待値

確率空間 $(\Omega, \mathcal{F}, P)$ の上に与えられた確率変数 X に対して：

- 確率変数の**期待値**の一般的定義：$E[X] \stackrel{\text{def}}{=} \int_\Omega X(\omega)P(d\omega)$．右辺は測度 P に関する一般の「積分」として定義できるが，ここでは深入りしない．

- 期待値は「積分」なので発散（∞ や $-\infty$）する場合もある．
- 以下で判るように，確率変数 X と Y が同じ分布に従うなら，$E[X] = E[Y]$．つまり，**期待値は分布で決まる**．

実際には，測度 P と確率変数 X が明示的（具体的）に与えられるわけではなく，X が従う分布関数 $F(x) = \Pr[X \le x] = P(\{\omega | X(\omega) \le x\})$ が与えられ，それに基づいて期待値を計算するので,「**確率変数の期待値**」は実は「(確率変数が従う）**分布の期待値**（次式の最右辺)」である．

- **期待値の計算**：$E[X] = \int_\Omega X(\omega) P(d\omega) = \int_{-\infty}^{\infty} xF(dx)$
  - 離散（確率関数 $\{p(k); k = 0, 1, \ldots\}$ の場合） $E[X] = \sum_{k=0}^{\infty} kp(k)$
  - 連続（密度関数 f の場合） $E[X] = \int_{-\infty}^{\infty} xf(x)dx$

最も単純な架空の例で，これらの意味を説明する．2 つのコインがあり，各コインの出目を表す（表が 1 で裏が 0）確率変数を X, Y と置く．神様が選択する世界の運命は 3 通りしかなく（$\{\omega_0, \omega_1, \omega_2\}$)，神様がどれか 1 つの運命を平等に（確率 1/3）選択し，運命とコインの出目の対応を以下の表とする．

運命	ω_0	ω_1	ω_2
発生確率	1/3	1/3	1/3
$X(\omega)$	0	1	1
$Y(\omega)$	1	1	0

本来の意味での期待値 $\int_\Omega X(\omega)P(d\omega)$ を，運命 ω 上の関数 (X や Y）の値の，その**運命の起きる確率を重みとした加重和**と解釈すれば（リーマン積分的に)，各々,

$$E[X] = 0 \times \frac{1}{3} + 1 \times \frac{1}{3} + 1 \times \frac{1}{3} = \frac{2}{3},\ E[Y] = 1 \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3} = \frac{2}{3}$$

となる（結果的に $E[X] = E[Y]$)．一方，これを，$\int_{-\infty}^{\infty} xF(dx) = \sum_{k=0}^{\infty} k \Pr[X = k]$，すなわち**「値」の出る確率（分布によって決まる）の加重和**に書きなおすと，X も Y も同じ分布であり，1 が出る確率は 2/3，0 が出る確率は 1/3 なので,

$$E[X] = E[Y] = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{2}{3} \qquad \text{（高校までの「期待値」の定義）}$$

となる．実は，$\int_{-\infty}^{\infty} xF(dx)$ はルベーグ積分であり,「値」の「その値を取る運命 ω の集合」の測度（確率）の加重和と解釈できるので，2 つが一致するのは当然である．なお，この例の X と Y は**独立ではない**（「確率変数の独立」: 第 6 講参照)．

- **基本性質：**
  - $\Pr[0 \leq X] = 1$ ならば，$0 \leq E[X]$.
    - 特に $\Pr[0 \leq X] = 1$ かつ $0 = E[X]$ ならば，$\Pr[X = 0] = 1$.
  - $E[aX + bY + c] = aE[X] + bE[Y] + c$　（X と Y が**独立でなくても**成立）．a, b, c は定数．証明はオンライン付録.

- 補足１：期待値（を定義する積分）が発散する例（コーシー分布）.
  密度関数が $f(x) = \dfrac{\lambda}{\pi(\lambda^2 + (x-\mu)^2)}$ の場合，期待値が発散する．

$$\int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{\infty} \frac{\lambda x}{\pi(\lambda^2 + (x-\mu)^2)} dx \ = \ \infty$$

  無限区間の積分で片側各々が発散するので全体として発散.

- 補足２：確率変数の「期待値」と観測データの「平均値」の関係.
  期待値のことを「平均（値）」と呼ぶ場合もある．しかし，確率変数の「平均値（＝期待値）」と観測データの「平均値（相加平均）」を混同しないことが重要である.
  - サイコロ X の**期待値** $E[X] = 1 \cdot \dfrac{1}{6} + \cdots + 6 \cdot \dfrac{1}{6} = 3.5$ は理論（確定）値.
  - サイコロ X を実際に 10 回投げた結果（運命が ω の時の観測データ）$\{X_1(\omega), X_2(\omega), \ldots, X_{10}(\omega)\}$ の（相加）**平均値**は，$\dfrac{1}{10}\displaystyle\sum_{i=1}^{10} X_i(\omega)$ で，実験の毎に異なる値であり，3.5 に近いことが多い．もちろん（10 回とも 1 が出て）1 になることも,（10 回とも 6 が出て）6 になることもありえる．運命 ω を引数として確率変数と見なせる.

    『統計的推定』では，観測データの「平均」を，観測されたデータから確率変数（分布）の「未知の期待値」を推定する方法と見なす（後述）.

条件付期待値

２つの（独立でない）確率変数 X, Y があり，$E[X] < \infty$ とする．Y の値 y が知られている場合の，X の「期待値」を**条件付期待値**と呼び，$E[X|Y = y]$ と書く．$Y = y$ の時の X の**条件付分布**の「期待値」である.

- (X, Y) が**結合確率関数** $\{p(j, k); j, k = 0, 1, \ldots\}$ を持つ場合（離散）：
  Y の（周辺）確率関数 $p_Y(y) = \displaystyle\sum_{x=0}^{\infty} p(x, y)$ より，$Y = y$ の時の X の条件付確率関数は $p_{X|Y}(x|y) = \dfrac{p(x, y)}{p_Y(y)}$ で（ただし $p_Y(y) > 0$ なら），$Y = y$ の時

の X の**条件付期待値**は,

$$E[X|Y=y] = \sum_{x=0}^{\infty} x p_{X|Y}(x|y) = \frac{1}{p_Y(y)} \sum_{x=0}^{\infty} x\, p(x,y) = \frac{\sum_{x=0}^{\infty} x\, p(x,y)}{\sum_{x=0}^{\infty} p(x,y)}$$

- (X,Y) が**結合密度関数** $f(x,y)$ を持つ場合（連続）：
  Y の（周辺）密度関数 $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$ より，$Y=y$ の時の X の条件付密度関数は $f_{X|Y}(x|y) = \dfrac{f(x,y)}{f_Y(y)}$ で（ただし $f_Y(y) > 0$ なら），$Y=y$ の時の X の**条件付期待値**は,

$$E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx = \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} x f(x,y)dx = \frac{\int_{-\infty}^{\infty} x\, f(x,y)dx}{\int_{-\infty}^{\infty} f(x,y)dx}$$

一般の条件付期待値は確率変数として定義される（→オンライン付録）.

<u>確率変数の分散と標準偏差</u>

X^2 の期待値が有限 ($E[X^2] < \infty$）の場合，X の「分散」が定義できる.

- **分散**：$V[X] \stackrel{\text{def}}{=} E[(X-E[X])^2]$
  $= E[X^2] - (E[X])^2$　（変形の計算を確かめよ）
  $= E[X(X-1)] + E[X](1-E[X])$　（変形の計算を確かめよ）

例：サイコロ（の出目）X の**分散**は,

$$V[X] = E[X^2] - (E[X])^2 = (1^2 \cdot \frac{1}{6} + \cdots + 6^2 \cdot \frac{1}{6}) - (3.5)^2 = \frac{35}{12}$$

- **基本性質**：X, Y は有限な分散を持つ確率変数とする.
  - $0 \le V[X]$.　$\sqrt{V[X]}$ を **<u>標準偏差</u>** と呼ぶ.
  - $V[aX+b] = a^2 V[X]$，(a, b は定数）
  - $V[X+Y] = V[X] + V[Y] + 2(E[XY] - E[X]E[Y])$
    $= V[X] + V[Y] + 2Cov[X,Y]$
    ここで，$(E[XY] - E[X]E[Y])$ を **<u>共分散</u>**$Cov[X,Y]$ と呼び，X と Y が**独立**ならば 0（後述.「逆」は成り立たない）.
  - よって，X と Y が**独立**ならば，$V[X+Y] = V[X] + V[Y]$.

- **変動係数**：X が**非負**の場合，$\delta[X] \stackrel{\text{def}}{=} \dfrac{\sqrt{V[X]}}{E[X]}$

  単位によらないばらつき度合いを示す（$\dfrac{\sqrt{V[aX]}}{E[aX]} = \dfrac{a\sqrt{V[X]}}{aE[X]} = \dfrac{\sqrt{V[X]}}{E[X]}$）.

- **標準化（正規化）**： $Y(\omega) \overset{\mathrm{def}}{=} \dfrac{X(\omega)-E[X]}{\sqrt{V[X]}}$. よって $E[Y]=0,\ V[Y]=1$
- k-**次モーメント**： $E[X^k]$．期待値は 1-次モーメント.
- k-**次中心モーメント**： $E[(X-E[X])^k]$．分散は 2-次中心モーメント.

重要 **確率変数の関数（合成確率変数）の期待値**

確率変数 X（分布 $F(x)$），実数値関数 $g(x)$ に対し，確率変数 $g(X)$ の期待値：

- $E[g(X)] = \int_\Omega g(X(\omega))P(d\omega) = \int_{-\infty}^{\infty} g(x)F(dx) = \int_{-\infty}^{\infty} g(x)f(x)dx$

ただし，最右辺は密度関数 f が存在する場合（→証明はオンライン付録）.

確率ベクトル $(X_1, X_2, \ldots, X_n)$ と実数値関数 $g(x_1, x_2, \ldots, x_n)$ の合成の場合も，確率ベクトルの結合分布関数を H，結合密度関数を h と置くと，同様に，

- $E[g(X_1,\ldots,X_n)] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\ldots,x_n)H(dx_1,\ldots,dx_n)$
  $= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(x_1,\ldots,x_n)h(x_1,\ldots,x_n)dx_1,\ldots,dx_n$

- 重要な系：**独立な確率変数の積の期待値＝期待値の積**

  互いに独立な確率変数 $X_1, X_2, \ldots, X_n$ において，どれもその絶対値が（有限の）期待値を持つ場合の，積の期待値：$E[\prod_i X_i] = \prod_i E[X_i]$

これは，$X_1, X_2, \ldots, X_n$ の各々の密度関数を $f_1(x), f_2(x), \ldots, f_n(x)$ とすると，結合密度関数は $h(x_1, x_2, \ldots, x_n) = f_1(x_1)\cdot f_2(x_2)\cdots f_n(x_n)$ なので，

$$\begin{aligned} E[X_1X_2\cdots X_n] &= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x_1x_2\cdots x_n f_1(x_1)f_2(x_2)\cdots f_n(x_n)dx_1dx_2\cdots dx_n \\ &= \prod_i\left(\int_{-\infty}^{\infty} x_if_i(x_i)dx_i\right) = \prod_i E[X_i] \end{aligned}$$

例題 **基本計算**

＜1＞ 確率変数 X と Y は独立．X, X^2, Y, Y^2 の期待値を各々，$m_1 = E[X]$，$m_2 = E[X^2]$，$M_1 = E[Y]$，$M_2 = E[Y^2]$ と置く（m_1, m_2, M_1, M_2 は有限値）．a, b, c は任意定数．この時，以下の値を $m_1, m_2, M_1, M_2, a, b, c$ で表現せよ.

- $E[-aX+bY+c] = -aE[X]+bE[Y]+c = -am_1+bM_1+c$
- $V[-aX+bY+c] = a^2V[X]+b^2V[Y] = a^2(m_2-m_1^2)+b^2(M_2-M_1^2)$

＜2＞ 実数区間 $[1,2]$ 上の一様分布に従う確率変数 X の期待値と分散を計算せよ.

$$E[X] = \int_1^2 xdx = \frac{1}{2}[x^2]_{x=1}^2 = \frac{3}{2},\ V[X] = E[X^2]-(E[X])^2 = \int_1^2 x^2dx - \frac{9}{4} = \frac{1}{12}$$

補分布による期待値の計算（→証明はオンライン付録）

X の従う分布関数を $F(x)$ と置く.

(1) **確率変数 X が非負値**の場合．期待値有限（$E[X] < \infty$）が前提.

連続型（X は連続型分布 F に従い，非負整数値を取る）：

$$E[X] = \int_0^\infty (1 - F(x))dx,$$

離散型（X が離散型分布 F に従い，非負整数値を取る）：　同様にして，

$$E[X] = \sum_{k=0}^{\infty} (1 - F(k)),$$

(2) **確率変数 X が負の値も取る**一般の場合．期待値有限（$E[X] < \infty$）が前提.

$$E[X] = \int_0^\infty (1 - F(x) - F(-x))\, dx$$

練習１

独立な確率変数 X, Y があり，共に実数区間 $[0, 1]$ 上の一様分布に従う場合の，差の絶対値（つまり確率変数 $|X - Y|$）の密度関数を求め,「期待値」を計算せよ.

練習２（小レポート課題３⇒次回提出）

独立な確率変数 X, Y があり，共に実数区間 $[1, 2]$ 上の一様分布に従う場合の，確率変数 $\dfrac{X}{Y}$ の「期待値」と「分散」を計算せよ.

ヒント： X, Y の各々の密度関数が与えられ，かつ X, Y が独立なので (X, Y) の結合密度関数が各密度関数の積で計算できる.

なお，この問題を解けば判るように,

$$E[\frac{X}{Y}] = E[X]E[\frac{1}{Y}] \geq \frac{E[X]}{E[Y]},$$

つまり，$E[\frac{1}{Y}] \geq \dfrac{1}{E[Y]}$ となるが，$f(x) = \dfrac{1}{x}$ は下に凸なので，一般に以下の**イェンゼンの不等式**の帰結である.

イェンゼンの不等式

区間 T 上に値を取る確率変数 X と T 内で下に凸な連続関数 $f(x)$ に対し，$E[f(X)] \geq f(E[X])$ が成り立つ（ただし $E[X], E[f(X)]$ が有限の場合）．証明はオンライン付録.

オンライン付録

A-1. 条件付期待値とラドン・ニコディムの定理

一般（厳密）には，条件付期待値は確率変数として定義される．本講義のレベルを超える問題を扱う際に必要になる．任意の実数 a に対して，

$$E[h(Y)I_{Y\le a}] \;=\; E[XI_{Y\le a}]$$

を満たす実数値関数 $h(x)$ が存在する時,

- **確率変数** $\underline{h(Y)}$ を $E[X|Y]$（X の Y に関する条件付期待値）,
- **関数** $\underline{h(y)}$ を $E[X|Y=y]$（X の $Y=y$ の時の条件付期待値）,

と定義する．ただし事象 A を示す確率変数を $I_A(\omega) \;=\; \begin{cases} 1 & \omega \in A \\ 0 & \text{otherwise} \end{cases}$ とする．確率変数 X, Y に対してこのような関数 h が存在し，ある意味で一意であることが「ラドン・ニコディムの定理」によって保証される．

逆に，条件付確率も， $P(A|B) = E[I_A|I_B = 1]$ で定義される．

A-2. 確率変数の関数の期待値の証明

X の密度関数 $f(x)$，実数関数 $h(x)$，$Y \stackrel{\text{def}}{=} h(X)$（ただし $E[Y] < \infty$）として，

$$E[Y] = \int_{-\infty}^{\infty} h(x)f(x)dx$$

（証明） $G(y) \stackrel{\text{def}}{=} \Pr[Y \le y] = \Pr[h(X) \le y] = \int_{h(x)\le y} f(x)dx = 1 - \int_{h(x)>y} f(x)dx$.
$H(x,y) \stackrel{\text{def}}{=} \begin{cases} 1 & h(x) > y \\ 0 & h(x) \le y \end{cases}$, $\int_0^{\infty} H(x,y)dy \;=\; \int_0^{\max(h(x),0)} dy \;=\; \begin{cases} h(x) & h(x) > 0 \\ 0 & \text{otherwise} \end{cases}$
を利用し，一般の確率変数の補分布と分布による期待値の表現を使うと,

$$\begin{aligned}
E[Y] \;&=\; \int_0^{\infty}(1 - G(y) - G(-y))dy = \int_0^{\infty}(1 - G(y))dy - \int_{-\infty}^{0} G(z)dz \\
&=\; \int_0^{\infty}\left(\int_{h(x)>y} f(x)dx\right)dy - \int_{-\infty}^{0}\left(\int_{h(x)\le z} f(x)dx\right)dz \\
&=\; \int_0^{\infty}\left(\int_{-\infty}^{\infty} f(x)H(x,y)dx\right)dy - \int_{-\infty}^{0}\left(\int_{-\infty}^{\infty} f(x)(1 - H(x,z)dx\right)dz \\
&=\; \int_{-\infty}^{\infty} f(x)\left(\int_0^{\infty} H(x,y)dy - \int_{-\infty}^{0}(1 - H(x,z))dz\right)dx \\
&=\; \int_{-\infty}^{\infty} f(x)(\max(h(x),0) - \max(-h(x),0))dx = \int_{-\infty}^{\infty} f(x)h(x)dx
\end{aligned}$$

最後の行の中の変形を詳しく書くと，

$$\int_0^\infty H(x,y)dy \;=\; \int_0^{h(x)} dy = h(x) \quad \text{if}\,(h(x)>0); \quad = \; 0 \quad (h(x)\le 0),$$

$$\int_{-\infty}^0 (1-H(x,y))dy \;=\; 0 \quad \text{if}\,(h(x)>0); \quad = \int_{h(x)}^0 dy = -h(x) \quad (h(x)\le 0),$$

$$\int_0^\infty H(x,y)dy - \int_{-\infty}^0 (1-H(x,y))dy \;=\; \begin{cases} h(x)-0 \;=\; h(x) & h(x)>0 \\ 0-(-h(x)) \;=\; h(x) & h(x)\le 0 \end{cases}$$

同様に多変数（ここでは２変数）の場合，X,Y の結合密度関数 $f(x,y)$，２変数実数関数 $h(x,y)$，$Z \stackrel{\text{def}}{=} h(X,Y)$（ただし $E[Z]<\infty$）として，

$$E[Z] = \int_{-\infty}^\infty \int_{-\infty}^\infty h(x,y)f(x,y)dxdy$$

（証明） $G(z) \stackrel{\text{def}}{=} \Pr[h(X,Y)\le z] = \int_{h(x,y)\le z} f(x,y)dxdy = 1 - \int_{h(x,y)>z} f(x,y)dxdy$.
$H(x,y,z) \stackrel{\text{def}}{=} \begin{cases} 1 & h(x,y)>z \\ 0 & h(x,y)\le z \end{cases}$, $\int_0^\infty H(x,y,z)dz \;=\; \begin{cases} h(x,y) & h(x,y)>0 \\ 0 & \text{otherwise} \end{cases}$ を利用し，一般の確率変数の補分布と分布による期待値の表現を使うと，

$$\begin{aligned} E[Z] \;&=\; \int_0^\infty (1-G(z)-G(-z))dz = \int_0^\infty (1-G(z))dz - \int_{-\infty}^0 G(w)dw \\ &=\; \int_0^\infty \left(\int_{h(x,y)>z} f(x,y)dxdy\right) dz - \int_{-\infty}^0 \left(\int_{h(x,y)\le w} f(x,y)dxdy\right) dw \\ &=\; \int_0^\infty \left(\int_{-\infty}^\infty \int_{-\infty}^\infty f(x,y)H(x,y,z)dxdy\right) dz \\ &\quad - \int_{-\infty}^0 \left(\int_{-\infty}^\infty \int_{-\infty}^\infty f(x,y)(1-H(x,y,w)dxdy\right) dw \\ &=\; \int_{-\infty}^\infty \int_{-\infty}^\infty f(x,y) \left(\int_0^\infty H(x,y,z)dz - \int_{-\infty}^0 (1-H(x,y,w))dw\right) dxdy \\ &=\; \int_{-\infty}^\infty \int_{-\infty}^\infty h(x,y)f(x,y)dxdy \end{aligned}$$

念のため，この「確率変数の関数の期待値」の計算を使って,「和の期待値」＝「期待値の和」という基本性質を示しておく.

X,Y の結合密度関数を $f(x,y)$，各々の（周辺）密度関数 $f_X(x), f_Y(y)$ と置くと，$f_X(x) = \int_{-\infty}^\infty f(x,y)dy,\ f_Y(y) = \int_{-\infty}^\infty f(x,y)dx$．そこで，

$$\begin{aligned} E[X+Y] \;&=\; \int_{-\infty}^\infty \int_{-\infty}^\infty (x+y)f(x,y)dxdy = \int_{-\infty}^\infty xf_X(x)dx + \int_{-\infty}^\infty yf_Y(y)dy \\ &=\; E[X]+E[Y] \end{aligned}$$

A-3. 補分布による期待値や分散の表現

(1) **確率変数 X が非負値**の場合．期待値有限（$E[X]<\infty$）が前提.

<u>連続型</u>（X は連続型分布 F に従い，非負整数値を取る）：

$$E[X]=\int_0^\infty(1-F(x))dx,\ E[X^2]=2\int_0^\infty x(1-F(x))dx,\ V[X]=E[X^2]-(E[X])^2$$

証明：密度関数を f．$\mathbf{I}(x,y)=\begin{cases}1 & x\geq y\\ 0 & x<y\end{cases}$ として，$x=\int_0^x dy$ も利用すると，

$$\begin{aligned}E[X] &= \int_0^\infty xf(x)dx=\int_0^\infty\left(\int_0^x dy\right)f(x)dx=\int_0^\infty\left(\int_0^\infty \mathbf{I}(x,y)dy\right)f(x)dx\\ &= \int_0^\infty\left(\int_0^\infty \mathbf{I}(x,y)f(x)dx\right)dy=\int_0^\infty\left(\int_y^\infty f(x)dx\right)dy=\int_0^\infty(1-F(y))\,dy\end{aligned}$$

$$E[X^2]=\int_0^\infty x^2f(x)dx=\int_0^\infty\left(\int_0^x 2ydy\right)f(x)dx=\cdots=2\int_0^\infty x(1-F(x))dx$$

<u>離散型</u>（X が離散型分布 F に従い，非負整数値を取る）： 同様にして，

$$E[X]=\sum_{k=1}^\infty(1-F(k-1)),\ E[X(X-1)]=\sum_{k=1}^\infty 2k(1-F(k)),$$

$$V[X]=E[X(X-1)]+E[X](1-E[X])$$

(2) **確率変数 X が負の値も取る**一般の場合．期待値有限（$E[X]<\infty$）が前提.

X の従う分布を F_X と置く．非負値部分と負値部分に分ける.

$$E[X]=E[X^+]-E[X^-]=\int_0^\infty(1-F_X(x)-F_X(-x))\,dx$$

$$X^+(\omega)=\begin{cases}X(\omega) & \text{if } X(\omega)\geq 0\\ 0 & \text{その他の}\omega\end{cases},\ X^-(\omega)=\begin{cases}-X(\omega) & \text{if } X(\omega)<0\\ 0 & \text{その他の}\omega\end{cases},$$

証明：ここで，$x>0$ では（正確にはほとんど至るところの $x>0$ で），

- $F_{X^+}(x)=\Pr[X^+\leq x]=\Pr[X\leq x]=F_X(x)$
- $F_{X^-}(x)=\Pr[X^-\leq x]=\Pr[-x\leq X]=1-\Pr[X<-x]=1-F_X(-x)$

であり，X^+ も X^- も有限な非負値確率変数なので，

- $E[X^+]=\int_0^\infty(1-F_{X^+}(x))\,dx=\int_0^\infty(1-F_X(x))\,dx$
- $E[X^-]=\int_0^\infty(1-F_{X^-}(x))\,dx=\int_0^\infty F_X(-x)dx$

A-4. モーメント母関数

モーメント母関数：$E[e^{tX}]$ を利用した期待値・分散の計算.

$M(t) \overset{\text{def}}{=} E[e^{tX}] = \begin{cases} \sum_{k=-\infty}^{\infty} e^{tk}p(k) & \text{（離散型：}p(x)\text{ は確率関数）} \\ \int_{-\infty}^{\infty} e^{tx}f(x)dx & \text{（連続型：}f(x)\text{ は密度関数）} \end{cases}$ が $t = 0$ の近くで発散しない場合,

- $\displaystyle\lim_{t\to 0}\frac{d}{dt}M(t) = E[X]$, $\displaystyle\lim_{t\to 0}\frac{d^2}{dt^2}M(t) = E[X^2]$, $V[X] = E[X^2] - (E[X])^2$

によって計算する方が簡単なこともある．他に，特性関数や確率母関数というものもある．

（計算）：X の密度関数を $f(x)$ として，$E[e^{tX}] = \displaystyle\int_{-\infty}^{\infty} e^{tx}f(x)dx$ より，$M(t)$ は

$$= \int_{-\infty}^{\infty} e^{tx}f(x)dx = \int_{-\infty}^{\infty}\sum_{n=0}^{\infty}\frac{t^n x^n}{n!}f(x)dx = \sum_{n=0}^{\infty}\frac{t^n}{n!}\int_{-\infty}^{\infty}x^n f(x)dx = \sum_{n=0}^{\infty}\frac{t^n}{n!}E[X^n]$$

よって，$\displaystyle\frac{d}{dt}M(t) = \sum_{n=1}^{\infty}\frac{nt^{n-1}}{n!}E[X^n] = E[X] + \sum_{m=1}^{\infty}\frac{t^m}{m!}E[X^{m+1}]$ （$m = n-1$）より，$\displaystyle\lim_{t\to 0}\frac{d}{dt}M(t) = E[X]$．以下同様にして $\displaystyle\frac{d^k}{dt^k}M(t) = E[X^k] + \sum_{n=1}^{\infty}\frac{t^n}{n!}E[X^{n+k}]$.

A-5. イェンゼンの不等式

区間T上に値を取る確率変数XとT内で下に凸な連続関数$f(x)$が，$E[X], E[f(X)] < \infty$ を満たす時，$E[f(X)] \geq f(E[X])$ が成り立つ．

（証明）：$X(\omega) \in T$ $(\forall\omega)$ なので，$x_0 = E[X]$ も $x_0 \in T$．T 内の任意の点 $x = x_0$ を固定し，その点で曲線$y = f(x)$ に接する直線のうちの任意の１つを$y = a_0(x - x_0) + f(x_0)$ とする．

f が下に凸 $\Leftrightarrow$ $f(x) \geq a_0(x - x_0) + f(x_0)$ $(\forall x \in T)$．すなわち，$f(X(\omega)) \geq a_0(X(\omega) - x_0) + f(x_0)$ $(\forall\omega)$．よって，

$$E[f(X)] \geq E[a_0(X - x_0) + f(x_0)] = a_0(E[X] - E[X]) + f(E[X]) = f(E[X])$$

例えば，$f(x) = \dfrac{1}{x}$ は下に凸なので，練習 2 の確率変数Y に対して適用すると，$E[\dfrac{1}{Y}] = \log 2 = 0.693\ldots > \dfrac{1}{E[Y]} = \dfrac{2}{3} = 0.666\ldots$ となる．

A-6. 複雑な分布の期待値

X が分布 $F(x) = \begin{cases} 1 - e^{-2x} & (x \geq 1) \\ 0 & (x < 1) \end{cases}$ に従う時の期待値 $E[e^X]$．これは補分布を用いて期待値を計算すると簡単になる例である．

分布が $x=1$ で不連続である点に注意．非負の確率変数 e^X の分布関数 $G(x) \stackrel{\text{def}}{=} \Pr[e^X \leq x]$ は， $G(x) = \begin{cases} \Pr[X \leq \log x] = 1 - e^{-2\log x} = 1 - x^{-2} & (x \geq e) \\ 0 & (x < e) \end{cases}$ で，

$$E[e^X] = \int_0^\infty (1 - G(x))dx = \int_0^e dx + \int_e^\infty x^{-2}dx = e + e^{-1}$$

別解：

上で見たように $F(x)$ が連続で密度関数 f を持つなら，任意の関数 g に対して，$g(X)$ の期待値 $E[g(X)]$ は， $E[g(X)] = \int_{-\infty}^\infty g(x)f(x)dx$ で計算できる．

分布 F が不連続な場合，上のように単純計算はできないが，F の不連続点を $x_1, x_2, \cdots$ とし，$F(x) = F_c(x) + F_s(x)$（F_c は区分的に微分可能な連続関数，F_s は階段関数）と分解できる場合は，$f(x) \stackrel{\text{def}}{=} F_c'(x)$ と置いて

$$E[g(X)] = \int_{-\infty}^\infty g(x)f(x)dx + \sum_{x_i} g(x_i)(F_s(x_i + 0) - F_s(x_i - 0))$$

よって本問の場合は，$f(x) = F_c'(x) = 2e^{-2x}\ (x \geq 1),\ g(x) = e^x,\ x_i = 1$ となり，

$$E[e^X] = \int_1^\infty 2e^x e^{-2x}dx + e(1 - e^{-2}) = 2\int_1^\infty e^{-x}dx + e - e^{-1} = e + e^{-1}$$