Lecture notes MVE136 Fall 2019. Part 2: Random Processes

8. Random processes

Definition 8.1. A *random process* is a family $X(t) = X(t, \xi)$ of r.v.'s indexed by time $t \in T$ in a time parameter set T.

So for each time $t \in T$ we have a random variable $X(t, \xi)$ which is a function of the outcome $\xi \in S$ of a random experiment. The time parameter set is either discrete, e.g., $T = \mathbb{N}$, $\mathbb{Z}$, $\{0, \ldots, n\}$ or continuous, e.g., $T = \mathbb{R}^+$, $\mathbb{R}$, $[a, b]$.

As the CDF (or PMF/PDF) is used to study r.v.'s one might belive the CDF $F_{X(t)}$ of a random process $X(t)$ for all $t \in T$ tell a lot about the process. This is not true:

Example 8.1. Consider the random processes $Y(t)$, $t \in \mathbb{R}$, and $\{Z(t)$,t$\in \mathbb{R}$, where each $Y(t)$ r.v. is $\mathrm{N}(0, 1)$ independent of all other process values $Y(s)$, $s \neq t$, while $Z(t) = X$ for the one and same $\mathrm{N}(0, 1)$ r.v. X for all t. Then $F_{Y(t)}(x) = F_{Z(t)}(x) = \Phi(x)$ for all t, x despite that $Y(t)$ and $Z(t)$ are more or less as different as they can be - the first totally independent and therefore widely oscillating - the other totally dependent equal to a single constant random value at all times.

To have full probabilistic information about a random process one need to know

$$F_{X(t_1),\ldots,X(t_n)}(x_1, \ldots, x_n) = \Pr(X(t_1) \leq x_1, \ldots, X(t_n) \leq x_n)$$

for all $t_1, \ldots, t_n \in T$, $x_1, \ldots, x_n \in \mathbb{R}$ and $n \in \mathbb{N}$.

Example 8.1. (Continued) For $Y(t)$ and $Z(t)$ as before we have

$$F_{Y(t_1),\ldots,Y(t_n)}(x_1, \ldots, x_n) = \Pr(Y(t_1) \leq x_1, \ldots, Y(t_n) \leq x_n) = \Phi(x_1) \cdot \ldots \cdot \Phi(x_n),$$

$$F_{Z(t_1),\ldots,Z(t_n)}(x_1, \ldots, x_n) = \Pr(Z(t_1) \leq x_1, \ldots, Z(t_n) \leq x_n) = \Phi(\min\{x_1 \ldots, x_n\}).$$

More partial information about the process that is often sufficient for needs in engineering mathematics is given by the following three functions:

Definition 8.2. 1. The *mean function* $\mu_X(t) = E(X(t))$ for $t \in T$,

2. the *autocorrelation function (ACF)* $R_{XX}(s, t) = E(X(s)X(t))$ for $s, t \in T$,
3. the *autocovariance function* $C_{XX}(s, t) = \mathrm{Cov}(X(s), X(t))$ for $s, t \in T$.

Note that $R_{XX}(s,t) = C_{XX}(s,t) + \mu_X(s)\mu_X(t)$.

Example 8.1. (Continued) Here $\mu_Y(t) = \mu_Z(t) = 0$, $R_{YY}(s,t) = C_{YY}(s,t) = 1$ and 0 for $s = t$ and $s \neq t$, respectively, while $R_{ZZ}(s,t) = C_{ZZ}(s,t) = 1$ for all s,t.

Example 8.2. (Cosine process) For $X(t) = U\cos(\omega t) + V\sin(\omega t)$ for $t \in \mathbb{R}$ where U and V are zero-mean uncorrelated r.v.'s with variance σ^2 and $\omega \in \mathbb{R}$ is a constant we have $\mu_X(t) = 0$ and

$$R_{XX}(s,t) = E(U^2)\cos(\omega s)\cos(\omega t) + E(V^2)\sin(\omega s)\sin(\omega t) = \sigma^2\cos(\omega(t-s)).$$

Example 8.3. For $X(t) = a\sin(\omega t + \Theta)$ for $t \in \mathbb{R}$ where Θ is uniformly distributed over $[0, 2\pi]$ and $\omega, a \in \mathbb{R}$ are constants we have $\mu_X(t) = 0$ and

$$R_{XX}(s,t) = a^2\,E\big(\tfrac{1}{2}\cos(\omega(s-t)) + \tfrac{1}{2}\cos(\omega(s+t)+\Theta)\big) = \tfrac{1}{2}\,a^2\cos(\omega(t-s))$$

as the mean of the second term in the middle step is zero by symmetry.

Definition 8.3. For two random processes $X(t)$ and $Y(t)$ we define

1. the *crosscorrelation function* $R_{XY}(s,t) = E(X(s)Y(t))$,
2. the *crosscovariance function* $C_{XY}(s,t) = \mathrm{Cov}(X(s), Y(t))$.

Definition 8.4. A random process $X(t)$ is *strictly stationary* if

$$F_{X(t_1+h),\ldots,X(t_n+h)}(x_1,\ldots,x_n) = F_{X(t_1),\ldots,X(t_n)}(x_1,\ldots,x_n) \text{ for all } t_1,\ldots,t_n \text{ and } h.$$

Strict stationarity means that the probability laws that governs the behaviour of the process are time invariant (but not that the process is constant or something such).

Example 8.1. (Continued) $Y(t)$ and $Z(t)$ as before are stationary as we have seen that their multivariate CDF's do not depend at all on the times involved.

Definition 8.5. A random process $X(t)$ is *wide sense stationary (WSS)* if $\mu_X(t) = \mu_X$ and $R_{XX}(t, t+\tau) = R_{XX}(\tau)$ do not depend on t.

Theorem 8.6. A strictly stationary process is WSS.

Proof. Follows from that $F_{X(t)}(x)$ and $F_{X(t),X(t+\tau)}(x,y)$ do not depend on t. □

All examples of process we have seen so far are WSS.

Theorem 8.7. (PROPERTIES OF WSS ACF) For $X(t)$ WSS we have

1. $R_{XX}(\tau) = R_{XX}(-\tau)$,
2. $E(X(t)^2) = R_{XX}(0)$,
3. $|R_{XX}(\tau)| \leq R_{XX}(0)$.

Proof. (1) $R_{XX}(\tau) = E(X(t)X(t+\tau)) = E(X(t+\tau)X(t)) = E(X(t)X(t-\tau)) = R_{XX}(-\tau)$.

(2) Take $\tau = 0$ in proof of (1).

(3) $0 \leq E\big((X(t) \pm X(t+\tau))^2\big) = 2\,R_{XX}(0) \pm 2\,R_{XX}(\tau)$. □

While $\cos(\tau)$ is an ACF by previous examples $\sin(\tau)$ is not by (1) and (3) above.

Definition 8.8. A WSS random processes $X(t)$, $t \in \mathbb{R}$, is

1. *ergodic in the mean* if $\lim\limits_{t\to\infty} \frac{1}{2t}\int_{-t}^{t} X(s)\,ds = \mu_X$,
2. *ergodic in the autocorrelation* if $\lim\limits_{t\to\infty} \frac{1}{2t}\int_{-t}^{t} X(s)X(s+\tau)\,ds = R_{XX}(\tau)$.

Theorem 8.9. A WSS process is ergodic in the mean if $\lim_{\tau\to\infty} R_{XX}(\tau) = \mu_X^2$.

Proof. As $E\big(\frac{1}{2t}\int_{-t}^{t} X(s)\,ds\big) = \mu_X$ it is enough to prove that

$$\mathrm{Var}\big(\tfrac{1}{2t}\textstyle\int_{-t}^{t}X(s)\,ds\big) = \mathrm{Cov}\big(\tfrac{1}{2t}\int_{-t}^{t}X(r)\,dr, \tfrac{1}{2t}\int_{-t}^{t}X(s)\,ds\big) = \tfrac{1}{4t^2}\int_{-t}^{t}\int_{-t}^{t}C_{XX}(r,s)\,drds \to 0$$

as $t \to \infty$ which holds as $C_{XX}(r,s) = R_{XX}(s-r) - \mu_X^2 \to 0$ for large $s-r$ and contributions to the integral for non-large $s-r$ are eventually nullified by factor $1/(4t^2)$. □

Example 8.3. (CONTINUED) The theorem do not apply to $X(t) = a\sin(\omega t + \Theta)$ but ergodicity in mean follows from that $\frac{1}{2t}\int_{-t}^{t} X(s)\,ds \in [-a\pi/(2t), a\pi/(2t)]$. Ergodicity in the autocorrelation follows from a version of the previous argument together with the fact that $X(s)X(s+\tau) = \frac{a^2}{2}\cos(\omega\tau) + \frac{a^2}{2}\cos(2\omega s + \omega\tau + \theta)$.

Definition 8.10. A random process $X(t)$, $t \in T$, is *Gaussian/normal* if $\sum_{i=1}^n a_i X(t_i)$ is a Gaussian r.v. for each choice of $t_1, \ldots, t_n \in T$ and $a_1, \ldots, a_n \in \mathbb{R}$.

Example 8.2. (CONTINUED) The process $X(t) = U\cos(\omega t) + V\sin(\omega t)$ with U and V independent $\mathrm{N}(0,\sigma^2)$ is Gaussian since

$$\begin{aligned}\textstyle\sum_{i=1}^n a_i X(t_i) &= U\textstyle\sum_{i=1}^n a_i\cos(\omega t_i) + V\sum_{i=1}^n a_i\sin(\omega t_i)\\ &= \mathrm{N}\Big(0, \sigma^2\big(\textstyle\sum_{i=1}^n a_i\cos(\omega t_i)\big)^2 + \sigma^2\big(\sum_{i=1}^n a_i\sin(\omega t_i)\big)^2\Big).\end{aligned}$$

Each Gaussian process value $X(t)$ is Gaussian but conversely the property that each process value is Gaussian is a much weaker demand than the process being Gaussian.

Theorem 8.11. A Gaussian process $X(t)$ is fully probabilistically determined by its mean function and autocovariance function (or ACF).

Proof. As $\sum_{i=1}^n \omega_i X(t_i)$ is $\mathrm{N}(m,\sigma^2)$ we have

$$\Phi_{X(t_1),\ldots,X(t_n)}(\omega_1,\ldots,\omega_n) = E\big(\mathrm{e}^{j\sum_{i=1}^n \omega_i X(t_i)}\big) = E\big(\mathrm{e}^{j\mathrm{N}(m,\sigma^2)}\big) = \mathrm{e}^{jm-\frac{1}{2}\sigma^2},$$

where $m = \sum_{i=1}^n \omega_i \mu_X(t_i)$ and $\sigma^2 = \sum_{i=1}^n \sum_{j=1}^n \omega_i\omega_j C_{XX}(t_i,t_j)$. □

Theorem 8.12. A Gaussian process is strictly stationary if WSS.

Proof. A WSS Gaussian process has same mean function and ACF as if strictly stationary since strict stationary implies WSS. But since the process is probabilistically determined by its mean function and ACF it is strictly stationary. □

Theorem 8.13. Two Gaussian process values are independent if uncorrelated.

Proof. A pair of uncorrelated process values have same means, variances and autocovariance as if independent but since their joint distribution is determined by their means, variances and autocovariance they are independent. □

Example 8.2. (CONTINUED) For U and V independent $\mathrm{N}(0,\sigma^2)$ as before we have $\Pr(X(1)+2\,X(2)>3) = \Pr(\mathrm{N}(m,\Sigma^2)>3) = 1-\Phi((3-m)/\Sigma)$ where $m = E(X(1)+2\,X(2)) = \mu_X + 2\,\mu_X = 0$ and $\Sigma^2 = \mathrm{Var}(X(1)+2\,X(2)) = \mathrm{Cov}(X(1)+2\,X(2), X(1)+2\,X(2)) = \mathrm{Var}(X(1))+4\,\mathrm{Cov}(X(1),X(2))+4\,\mathrm{Var}(X(2)) =_{\text{[zero-mean]}} R_{XX}(0)+4\,R_{XX}(1)+4\,R_{XX}(0) = 5\,\sigma^2+4\,\sigma^2\cos(\omega t)$.

There are two equivalent definitions of a Poisson process that occurs naturally as the counting process of the number of radioactive decays of a piece of radioactive matter:

Definition 8.14. A *Poisson process* $X(t)$, $t\geq 0$, with intensity $\lambda>0$ starts at $X(0)=0$, spends a exponentially distributed time with mean $1/\lambda$ at that value, then changes value to 1, spends an independent exponentially distributed time with mean $1/\lambda$ at that value, then changes value to 2, spends an independent exponentially distributed time with mean $1/\lambda$ at that value, then changes ... .

Definition 8.15. A *Poisson process* $X(t)$, $t \geq 0$, with intensity $\lambda > 0$ is given by

1. $X(0) = 0$,
2. $X(t+s) - X(s)$ is $\mathrm{Po}(\lambda t)$ for $s, t \geq 0$,
3. $X(t+s) - X(s)$ is independent of $X(r)$, $r \in [0, s]$, for $s, t \geq 0$.

Theorem 8.16. A Poisson process has $\mu_X(t) = \lambda\, t$ and $C_{XX}(s,t) = \lambda \min\{s,t\}$.

So a Poisson process is not WSS.

Proof. $\mu_X(t) = E(X(t)) = E(X(t) - X(0)) = E(\mathrm{Po}(\lambda t)) = \ldots = \lambda t$ while $C_{XX}(s,t) = \mathrm{Cov}(X(s), X(t)) = \mathrm{Cov}(X(s), X(s)) + \mathrm{Cov}(X(s), X(t) - X(s)) = \mathrm{Var}(X(s)) + 0 = \ldots = \lambda s$ for $s \leq t$ as $X(s)$ and $X(t) - X(s)$ are independent giving $C_{XX}(s,t) = \lambda \min\{s,t\}$. $\square$

Example 8.4. For a Poisson process $X(t)$ we have

$$\begin{aligned}\Pr(X(1) = 1 | X(2) = 2) = \tfrac{\Pr(X(1)=1,\, X(2)=2)}{\Pr(X(2)=2)} &= \tfrac{\Pr(X(1)=1,\, X(2)-X(1)=1)}{\Pr(X(2)=2)} \\ &= \tfrac{\Pr(X(1)=1)\,\Pr(X(2)-X(1)=1)}{\Pr(X(2)=2)} \\ &= \tfrac{\Pr(\mathrm{Po}(\lambda)=1)\,\Pr(\mathrm{Po}(\lambda)=1)}{\Pr(\mathrm{Po}(2\lambda)=2)} = \ldots = \tfrac{1}{2}.\end{aligned}$$

9. Markov chains

Definition 9.1. A discrete time discrete valued random process X_n, $n \in \mathbb{N}$, is called a *Markov chain* if

$$\Pr(X_{n+1} = j | X_n = i, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = \Pr(X_{n+1} = j | X_n = i) = p_{ij}$$

does not depend on $x_0, \ldots, x_{n-1}$ and n but only on i, j.

The probabilities p_{ij} are called *transistion probabilities* and are gathered in the *transition matrix* P with elements $(P)_{ij} = p_{ij}$.

Example 9.1. (Kid collecting super heroes) A fast food restaurant randomly gives kid one out of four different super hero figures at each visit of restaurant. The number X_n of different super heroes collected after n visits to restaurant is Markov chain with $X_0 = 0$, $p_{ii} = 0$, $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, 1 for $i = 0,1,2,3,4$ and $p_{i,i+1} = 1$, $\frac{3}{4}$, $\frac{1}{2}$, $\frac{1}{4}$ for $i = 0,1,2,3$ while $p_{ij} = 0$ for all other combinations of i, j.

Example 9.2. (SIMPLE RANDOM WALK) $X_0 = 0$ and $X_n = \sum_{i=1}^n Y_i$ for $n \geq 1$ where $\{Y_i\}_{i=1}^\infty$ are independent with $\Pr(Y_i = 1) = p$ and $\Pr(Y_i = -1) = 1 - p = q$ giving $p_{i,i+1} = p$ and $p_{i,i-1} = q$ while $p_{ij} = 0$ for all other combinations of i, j.

Example 9.3. (GAMBLERS RUIN) Gambler has initial fortune \$ $X_0 = d$ while "the house" has intial fortune \$ $b - d$. The fortune of the gambler X_n after n bets is a Markov chain with $p_{i,i+1} = p$ and $p_{i,i-1} = 1 - p = q$ for $i = 1, \ldots, b-1$ while $p_{00} = p_{bb} = 1$ and $p_{ij} = 0$ for all other combinations of i, j, i.e., gambler wins \$ one with probability p and loses \$ one with probability q in each bet but stops betting when $X_n = 0$ or b as then he or bank is broke, respectively.

Definition 9.2. The *n-step transition probabilities* $p_{ij}^{(n)} = \Pr(X_{m+n} = j | X_m = i)$ are elements of the *n-step transition matrix* $P^{(n)}$, i.e., $(P^{(n)})_{ij} = p_{ij}^{(n)}$.

Theorem 9.3. (CHAPMAN-KOLMOGOROV) $P^{(n)} = P^n$ for $n \in \mathbb{N}$.

Proof. Follows from induction using that

$$\begin{aligned}
(P^{(n+1)})_{ij} = \Pr(X_{m+n+1}{=}j | X_m{=}i) &= \tfrac{\Pr(X_{m+n+1}=j,\, X_m=i)}{\Pr(X_m=i)} \\
&= \sum_{\text{all } k} \tfrac{\Pr(X_{m+n+1}=j,\, X_{m+1}=k,\, X_m=i)}{\Pr(X_{m+1}=k,\, X_m=i)} \, \tfrac{\Pr(X_{m+1}=k,\, X_m=i)}{\Pr(X_m=i)} \\
&= \sum_{\text{all } k} \Pr(X_{m+n+1}{=}j | X_{m+1}{=}k, X_m{=}i) \, p_{ik} \\
&= \sum_{\text{all } k} (P^{(n)})_{kj} \, p_{ik} = (P \, P^{(n)})_{ij}. \qquad \square
\end{aligned}$$

Definition 9.4. The *distribution at time* n of Markov chain X_n is the row matrix $\pi(n)$ with elements $(\pi(n))_i = \Pr(X_n = i)$.

Theorem 9.5. $\pi(m + n) = \pi(m) \, P^n$ for $m, n \in \mathbb{N}$.

Proof. By the law of total probability and Chapman-Kolmorgorov

$$(\pi(m{+}n))_i = \Pr(X_{m+n}{=}i) = \sum_{\text{all } k} \Pr(X_{m+n}{=}i | X_m{=}k) \Pr(X_m{=}k) = \sum_{\text{all } k} (P^n)_{ki} \, (\pi(m))_k. \ \square$$

Definition 9.6. A row matrix π with positive elements is a *stationary distribution* for a Markov chain if $\pi \, P = \pi$ and $\sum_{\text{all } i} (\pi)_i = 1$.

Theorem 9.7. If $\pi(m) = \pi$ then $\pi(m + n) = \pi$ for $m, n \in \mathbb{N}$.

Proof. $\pi(m+n) = \pi(m)\, P^n = \pi\, P^n = \pi\, P^{n-1} = \ldots = \pi$. □

Example 9.1. (CONTINUED) For the super hero count chain $\pi = (0\ \ 0\ \ 0\ \ 0\ \ 1)$.

Definition 9.8. The *mean time to return* μ_i to a state (=possible value) i of a Markov chain X_n is defined $\mu_i = E(\min\{n \geq 1 : X_n = i\} | X_0 = i)$.

Definition 9.9. A state j is *accesible* from a state i ($i \to j$) if $p_{ij}^{(n)} > 0$ for some n while i and j *communicate* ($i \leftrightarrow j$) if $i \to j$ and $j \to i$.

If all states communicate the chain is called *irreducible*.

Simple random walk is irreducible but super hero count and gamblers ruin are not.

Definition 9.10. The *period* $d(i)$ of state i is defined $d(i) = \gcd\{n \geq 1 : p_{ii}^{(n)} > 0\}$.

The chain is *aperiodic* if all $d(i) = 1$.

Simple random walk has all $d(i) = 2$ while super hero count has $d(i) = 1$ for $i \geq 1$. The following theorem is very difficult to prove but is intiutively predictable:

Theorem 9.11. An irreducible aperiodic chain has a stationary distribution π if and only if all mean return times $\mu_i < \infty$ and in that case $(\pi)_i = 1/\mu_i$ for all i. Further, $p_{ij}^{(n)} \to 1/\mu_j$ as $n \to \infty$ for all j regardless of the μ_j's are finite or not.

Definition 9.12. State i is *transient* if $\Pr(X_n = i$ for some $n \geq 1 | X_0 = i) < 1$.

State i is *reccurent* if $\Pr(X_n = i$ for some $n \geq 1 | X_0 = i) = 1$.

Here are another two very important but quite difficult to prove theorems:

Theorem 9.13. For an irreducible chain either all states are reccurent or all states are transient. Further, all states have the same period and the mean return time $\mu_i < \infty$ for one state i if and only if $\mu_i < \infty$ for all states i

Theorem 9.14. State i is reccurent if and only if $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.

Example 9.2. (CONTINUED) For simple random walk we have $p_{ii}^{(n)} = 0$ and $\binom{n}{n/2} p^{n/2} (1-p)^{n/2}$ for n odd and even, respectively, as going from i to i in n steps means equally many steps $n/2$ up and down, so $p_{ii}^{(n)} = \Pr(\text{Bin}(n,p) = n/2)$.

By Stirlings's formula we have $n! \sim \sqrt{2\pi n}\, n^n \mathrm{e}^{-n}$ as $n \to \infty$ (in the sense that the ratio between the left hand side and right hand side of $\sim$ goes to 1 as $n \to \infty$).

To check whether simple random walk is reccurent or transient we check whether $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \sum_{n=1}^{\infty} \binom{2n}{n} p^n (1-p)^n = \infty$ or $< \infty$. By Striling's formula

$$\binom{2n}{n} p^n (1-p)^n = \frac{(2n)!\, p^n (1-p)^n}{(n!)\,(n!)} \sim \frac{\sqrt{4\pi n}\,(2n)^{2n} \mathrm{e}^{-2n} p^n (1-p)^n}{\sqrt{2\pi n}\, n^n \mathrm{e}^{-n} \sqrt{2\pi n}\, n^n \mathrm{e}^{-n}} = \frac{(4p(1-p))^n}{\sqrt{\pi n}} \quad \text{as } n \to \infty.$$

As $4p(1-p) = 1$ and < 1 for $p = \frac{1}{2}$ and $\neq \frac{1}{2}$ it follows from theorem above that chain is reccurent for $p = \frac{1}{2}$ and transient otherwise.

Example 9.1. (Continued) If $T = \min\{n \geq 1 : X_n = 4\}$ for super hero count chain we have by direct calculation

$$\begin{aligned} E(T) &= E\binom{\text{waiting time}}{\text{r.v. with } p=1} + E\binom{\text{waiting time}}{\text{r.v. with } p=\frac{3}{4}} + E\binom{\text{waiting time}}{\text{r.v. with } p=\frac{1}{2}} + E\binom{\text{waiting time}}{\text{r.v. with } p=\frac{1}{4}} \\ &= 1 + \tfrac{4}{3} + 2 + 4 = \tfrac{25}{3} \end{aligned}$$

as the consecutive moves from $0 \curvearrowright 1$, $1 \curvearrowright 2$, $2 \curvearrowright 3$ and $3 \curvearrowright 4$ probabilistically are equivalent to tossing unbalanced coin until heads with probability 1, $\frac{3}{4}$, $\frac{1}{2}$ and $\frac{1}{4}$ for heads, respectively, until first heads.

Consider modified super hero chain $\hat{X}_n$, $n \in \mathbb{N}$, where we change row five in the transition matrix from $(0\ 0\ 0\ 0\ 1)$ to $(1\ 0\ 0\ 0\ 0)$, that is $\hat{p}_{40} = 1$ for modified chain instead of $p_{44} = 1$ for original chain. Modfied chain $\hat{X}_n$ will be irreducible and aperiodic by inspection. Further, we have $E(T)$ for original chain equals $\hat{\mu}_0 - 1$ for modified chain, also by inspection. Solving the equation for stationary distribution $\hat{\pi}\,\hat{P} = \hat{\pi}$ and $\sum_{\text{all } i} (\hat{\pi})_i = 1$ we readily obtain $\hat{\pi} = \left(\frac{3}{28}\ \frac{4}{28}\ \frac{6}{28}\ \frac{12}{28}\ \frac{3}{28}\right)$. And we find again that $E(T) = \hat{\mu}_0 - 1 = 1/\hat{\pi}_0 - 1 = \frac{28}{3} = \frac{25}{3}$ using theorem above.

That mean $E(p)$ of waiting time r.v. X is $1/p$ can be seen from that $E(p) = 1 + (1-p)\,E(p)$ (as after one coin toss the probability is $1-p$ that you still are waiting for heads) or by differentiating the CHF $\Phi_X(\omega) = \ldots = p/(\mathrm{e}^{-j\omega} - (1-p))$ at zero.